

SEARCH REQUEST FORM

Requestor's
Name: _____Serial
Number: _____

Date: _____

Phone: _____

Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

STAFF USE ONLY

Date completed: 06-05-03
Searcher: Beverly e4994
Terminal time: 20
Elapsed time: _____
CPU time: _____
Total time: 25
Number of Searches: _____
Number of Databases: 1

Search Site

_____ STIC
_____ CM-1
_____ Pre-S

Type of Search

_____ N.A. Sequence
_____ A.A. Sequence
_____ Structure
_____ Bibliographic

Vendors

_____ IG Suite
_____ STN
_____ Dialog
_____ APS
_____ Geninfo
_____ SDC
_____ DARC/Questel
✓ Other CGN

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2003, 17:30:47 ; Search time 2758 Seconds
(without alignments)
10562.693 Million cell updates/sec

Title: US-09-482-788-1_COPY_4000_5000
Perfect score: 1001
Sequence: 1 caattgatgtctgcagcgc.....cccaagtctctgttacc 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pi:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pi:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rtd:*
- 36: em_htg_nam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	481.1	48.1	10934	8	FSESYNIA	Z18755 Fusarium sc
2	392	39.2	9633	6	BD013055	BD013055 Cyclic de
3	392	39.2	9633	23	BD010088	BD010088 Cyclic de
4	360	36.0	3155	8	COPEPSYNH	X96559 C.Oligosper
5	339.4	33.9	46899	6	A36768	A36768 Sequence 1
6	339.4	33.9	46899	8	AR050554	AR050554 Sequence
7	339.4	33.9	46899	6	TISACYS	Z28383 T.niveum (A
8	168.6	16.8	3973	6	A40406	A40406 Sequence 33
9	106.2	10.6	40989	1	AF204805	AF204805 Nostoc sp
10	97.6	9.8	39436	1	ASP269505	AJ269505 Anabaena
11	87.8	8.8	15582	1	AF455810	AF455810 Xenorhabd
12	86.2	8.6	3671	1	PSPBSC	X77699 Pseudomonas
13	86	8.6	28587	1	AF047828	AF047828 Pseudomon
14	85.6	8.6	37856	1	AX024319	AX024319 Sequence
15	85.6	8.6	37856	6	AX024212	AX024212 Sequence
16	84.8	8.5	216050	1	AL646076	AL646076 Ralstonia
17	81	8.1	4771	1	AF021263	AF021263 Streptomy
18	78.4	7.8	48774	1	AF007865	AF007865 Bacillus
19	77.6	7.8	7836	1	AF047717	AF047717 Streptomy
20	77.6	7.8	12744	1	AF204401	AF204401 Streptomy
21	77.4	7.7	18023	1	AX024392	AX024392 Sequence
22	77.4	7.7	18023	6	AX024285	AX024285 Sequence
23	77.2	7.7	3975	1	AMPEPSYNT	X97860 Amycolatops
24	75.6	7.6	22444	1	AE004669	AE004669 Pseudomon
25	75.6	7.6	31220	1	AE008316	AE008316 Agrobacte
26	75.4	7.5	7335	6	AX097467	AX097467 Sequence
27	75.4	7.5	15719	1	AE004667	AE004667 Pseudomon
28	74.8	7.5	23451	1	AB070952	AB070952 Streptomy
29	74.4	7.4	10569	1	AE009298	AE009298 Agrobacte
30	74.2	7.4	77457	1	AF210249	AF210249 Streptomy
31	73.6	7.4	28798	1	U95370	U95370 Bacillus li
32	73.2	7.3	28732	1	AF082100	AF082100 Streptomy
33	73	7.3	28890	1	AF512431	AF512431 Saccharot
34	73	7.3	37200	1	SC563	AL035640 Streptomy
35	72.4	7.2	7493	1	PP0310530	AJ310530 Pseudomon
36	72.4	7.2	32748	1	AB070951	AB070951 Streptomy
37	72	7.2	13964	1	BLAACHAA	AB070951 Streptomy
38	72	7.2	32386	1	BLAJ5061	AJ005061 Bacillus
39	71.8	7.2	3471	6	AX433752	AX433752 Sequence
40	71.6	7.2	2792	1	TSF318786	AJ318786 Tolypotr
41	70.4	7.0	5451	6	AX078430	AX078430 Sequence
42	70.4	7.0	8992	1	SV117268	Y17268 Streptomyce
43	70	7.0	7914	1	AF021262	AF021262 Streptomy
44	69.8	7.0	333500	1	AP003590	AP003590 Nostoc sp
45	69.2	6.9	230573	1	AF322012S2	AF322013 Bradyrhiz

ALIGNMENTS

RESULT 1
FSESYNIA
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

FSESYNIA
Fusarium scirpi esynl gene for enniatin synthetase.
Z18755
Z18755.3 GI:15591913
enniatiin synthetase; esynl gene.
Fusarium equiseti.
Fusarium equiseti
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreales; mitosporic Hypocreales; Fusarium.
1 (bases 1 to 10934)
Haese, A., Schubert, M., Hermann, M. and Zocher, R.
Molecular characterization of the enniatin synthetase gene encoding
a multifunctional enzyme catalysing N-methyldeipeptide formation

10934 bp DNA linear PLN 11-SEP-2001


```
Qy 595 GAGATCTGTCAGACACACCTCTCGCACAACTCATTTTGGAGTGCAGCTACAGAGGACC 654
||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5327 GAGATCTCTCGGAACACCGCTGGCACAGATAATGTTTGGCTTTCACCTCGCAGAGGACC 5386
Qy 655 TTGGAAGATTCAAGTTCAGGGTCTCGAGTCCGCTGACCTGTCCTGAGCAAGCGTACACTC 714
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5387 TTGGACGTTTTCGAGCTGGAGGTATCCAGCTGAGCCTATCGCCAGCAAGCGCTACACCA 5446
Qy 715 GATTTGACATGGAGTTCATCTGTTTCAAGAAACCGACAGCCTTAAAGGTAGCGTCAACT 774
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5447 GATTTCGATGTCGAGTTCATCTGTTTCAACAGGACAGCGGACTGAAGGCGAGTTGCCAACT 5506
Qy 775 TTGCCGATGAGCTGTTCAAAATGGAGACTGTTTGAATAATGTCGTCAGAGTATTTCTTGAGA 834
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5507 TTGCCACAGATCTCTTCAAGCCCGAAACTATCCAGAAATGTTGTCCAGCGTGTTCGCCAGA 5566
Qy 835 TTCTGAGAAAGCGGCTTCAAGTTTCGGGACACAGCTCTCAATACTTCTTCTGAGTATG 894
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5567 TTCTACGCCATGGGCTTTCACAGCCTGAGAGCTGTATCTCGGTTCTTCCACTGACTGATG 5626
Qy 895 GATTGTGACTCTTGAATAATTTGATGTTCTCAACGTCAAACATGTCGACTATCCCGAG 954
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5627 GAGTCGAGGACTTCGCGAGTTGGATCTGCTGGAAATCAAGAGGACTTAACCTCCGCG 5686
Qy 955 AATCGAGCTTGGCTGATGTTTCCAGACCCCAAGTCTCTGCTTACCC 1000
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5687 ATTCCAGCGTGTAGATGCTCTTCGCGAACAAGCTGCGGCCAACCC 5732

RESULT 2
BD013055 9633 bp DNA linear PAT 02-AUG-2002
LOCUS Cyclic depsipeptide synthetase and its gene and mass production
DEFINITION system of cyclic depsipeptide.
ACCESSION BD013055
VERSION BD013055.1 GI:22093244
KEYWORDS WO 0118179-A/1.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 9633)
AUTHORS Mido,N., Okakura,K., Miyamoto,K., Watanabe,M., Yanai,K.,
Yasutake,T., Aihara,S., Futamura,T., Kleinkauf,H. and Murakami,T.
TITLE Cyclic depsipeptide synthetase and its gene and mass production
JOURNAL system of cyclic depsipeptide
PATENT: WO 0118179-A 1 15-MAR-2001;
MEIJI SEIKA KAISHA LTD,NAOKI MIDO,KAORU OKAKURA,KOICHI MIYAMOTO,
MANABU WATANABE,KOJI YANAI,TETSUYA YASUTAKE,SATO AIHARA,AKAFUMI,
FUTAMURA, HORST KLEINKAUF, TAKESHI MURAKAMI
COMMENT OS Mycelia sterilia
PN WO 0118179-A/1
PD 15-MAR-2001
PF 07-SEP-2000 WO 2000JP006103
PR 07-SEP-1999 JP 99P 253040,06-APR-2000 JP 00P 104291 PI
NAOKI MIDO,KAORU OKAKURA,KOICHI MIYAMOTO,MANABU WATANABE, PI KOJI
YANAI,
PI TETSUYA YASUTAKE,SATO AIHARA,TAKAFUMI FUTAMURA,HORST
KLEINKAUF.
PI TAKESHI MURAKAMI
PC C12N9/00,C12N15/52,C12N1/15,C12P21/04
CC peptide synthetase for PF1022
FH Key Location/Qualifiers
FT CDS (1)..(9633)
FT mat_peptide (13)..(9630).
FEATURES Location/Qualifiers
1..9633
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 2318 a 2834 c 2462 g 2019 t
ORIGIN
```

Query Match

39.2%; Score 392; DB 6; Length 9633;

RESULT 3

```
Best Local Similarity 63.3%; Pred. No. 1.6e-102;
Matches 637; Conservative 0; Mismatches 360; Indels 9; Gaps 2;
Qy 1 CAATTGATGTTCTTGGAGCGCATCTCAATAGCTTACTAGCTCGCTCAAGGACTCAA 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3710 CTATCGACATCTAGCGCGGAACATAAGCAATTTCTATTACGCGCTCTCCGGGGCTC-- 3767
Qy 61 AAGACCCCTGTTCAGCACTCACTCTCTACCTATCCAGTACAGCGACTTTTGCAAATGGC 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3768 -TGATCTCTATCGTGTGAGCCCACTCCCACTCCACTACCGGACTTTTCCGTTTGGC 3826
Qy 121 AGAAG-----GACCAATTCATAGAGCAGGAGAGCACTCACTACTTGGAGAGAGCAAC 174
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3827 AAAAGCAGCTCGAACAGGAGACCGAACATGAGCGGCACTCGCAATACTGGGTCAAGCAGC 3886
Qy 175 TCAAGACTCTTCCCGACCAAGATCCCGACCGACTTTTCCCGCGCTTCGACTTCTGTCTG 234
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3887 TCGCAGACAGCTCGCGCGCCGCAATTCCTTAACGACTTCCCGCGCCCAACATACTGTCTG 3946
Qy 235 GAGACGAGTGTGCTGATACATGTTACATCGAGCGGAGCTCTTACAGTCCCTTCGAGCCT 294
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3947 GTGAAGCAGGTTCCGTCGACGTGAGATCGAAGGCGAACTGTATGAAGGCTCCAAGAT 4006
Qy 295 TCTGCAACGACACAAACAGGACTCTTCTGTCGTTCTTCTAGCTCGTTCGTCGCGCTC 354
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4007 TCTGTAAAGTAGAGCAAAATGACGCTTTTCCGCGTGTGTTAGGGGCTTTCGCGCGACCC 4066
Qy 355 ATTATCGTCTCACAGCTGTGTAAGAGCTGTGTTGTTGATACCAATTCGGAATCGCAACC 414
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4067 ATTATCGTCTCACCGCGCCGCAAGACTCGATCATCGGACGCCCATCGGAAACCCCAACC 4126
Qy 415 GACCTGAACGGAGGATATCATCGGCTGCTTTGTCAATACGAGTGTATCGAATCAACA 474
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4127 GCCAGGACTTGAACACATGATCGGCTTCTTCTGTCACACACCAATGATCGCAATCACGG 4186
Qy 475 TAGATCATCAGATACCTTTTGGGACTTTGATCAACAAGTCAAGGCTAGGACGACAGAG 534
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4187 TCGAGCGGACGACACTTTTGAAGCGCTGGTGGACAAAGTTCGGACCAAGGCGAGCGCG 4246
Qy 535 CATTCGAGAACGAGGATATTCGTTTGGAGCGGTGTATACAGCACTACAGCTGATGATCA 594
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4247 CATTCGAGCAACCAAGAGCTGCCCTTTGAGCGGCTCGTACGCGACTCTTCCACGCTCGA 4306
Qy 595 GAGATCTGTCAAGCAGACTCTCTCGACAACTCATTTTTCAGTGCAGCTCACAGAGGACC 654
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4307 GAGACTTATCCGAAACCCACTAGCACAGCTCACCTTCGCTCTTCTTCTCAACAGGACC 4366
Qy 655 TTGGAAGATTCAAGTTCAGGCTCTCGAGTCCGTCGAGTCCGTGCTAGCAAGCGTACACTC 714
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4367 TCGCAAGTTCGAGCTGGAGGCTCTGTAAGCGGAAACCCGCTCTCGAACAAGGTATACACA 4426
Qy 715 GATTTGACATGGAGTTCATCTGTTTCAAGAAACCGACAGCTTAAAGGTAGCGTCAACT 774
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4427 GGTTCGAGCTGGAGTTTCACTGTTTCCAGAAAGCCGGAAGACTAAGCGGTACGTTGGCAT 4486
Qy 775 TTCCGATGAGCTGTTTCAAAATGGAGACTGTTTGAATAATGTCGTCAGAGTATTTCTTTAGA 834
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4487 TTGCGCAGATCTATTCAAGCCTGAGACCATTAGCAATGTAGTCGCTATTTTTCCTCAAA 4546
Qy 835 TTCTCAGAAACGGGCTTCAAGTTTCGGGACACCAAGTCTCAATACTTCTTTTTCAGTGTG 894
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4547 TCCTCGCACAAAGGCATTCGCCAGCCTCGGACTCCAATCGCTGTTCTCCGCGTTACCGATG 4606
Qy 895 GCATTGTGACTCTTGAATAATTTGGATGTTTCAAGCTCAAACTGTCGACTATCCCGGAG 954
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4607 GGTTCAGCGACCTTCGTGCCATGGGCTTCTGATATCGAATCGAAGGAGCAATACCCCGGG 4666
Qy 955 AATCGAGCTTGGCTGATGTTTCCAGACCCCAAGTCTCTGCTTACCC 1000
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4667 AGTCGACGCTCGTCGACGCTCTTCCGCAAGCAGGTGGCGCGCTCACCC 4712
```

BD010088 ID BD010088 standard; DNA; UNC; 9633 BP.

XX AC BD010088;

XX SV BD010088.1

XX 08-FEB-2002 (Rel. 70, Created)

XX 08-FEB-2002 (Rel. 70, Last updated, Version 1)

XX Cyclic depsipeptide synthetase and its gene and mass production system of cyclic depsipeptide.

XX JP 03075848-T/1.

XX unidentified

XX unclassified.

XX [1]

XX 1-9633

XX Aihara S., Okakura K., Miyamoto K., Watanabe M., Yanai K., Yasutake T., Mido N., Futamura T., Kleinkauf H., Murakami T.;

XX "Cyclic depsipeptide synthetase and its gene and mass production system of cyclic depsipeptide";

XX Patent number JP03075848-T/1, 06-MAR-2001.

XX MELJI SEIKA KAISHA LTD.

XX OS Mycelia sterilia

XX PN JP 03075848-T/1

XX PD 06-MAR-2001

XX PF 07-SEP-2000 JP 200006103

XX PR 07-SEP-1999 JP 99P 253040, 06-APR-2000 JP OOP 104291

XX PI NAOKI MIDO, KAOBU OKAKURA, KOICHI MIYAMOTO, MANABU WATANABE, KOJI YANAI,

XX PI TETSUYA YASUTAKE, SATO AIHARA, TAKAFUMI FUTAMURA, HORST KLEINKAUF,

XX PI TAKESHI MURAKAMI

XX PC C12N9/00, C12N15/52, C12N1/15, C12P21/04

XX CC Key Location/Qualifiers

XX FH CDS (1)..(9633)

XX FT mat_peptide (13)..(9630)

XX Key Location/Qualifiers

XX source 1..9633

XX /db_xref="taxon:32644"

XX /organism="unidentified"

XX SQ Sequence 9633 BP; 2318 A; 2834 C; 2462 G; 2019 T; 0 other;

Query Match 39.2%; Score 392; DB 23; Length 9633;

Best Local Similarity 63.3%; Pred. No. 1.6e-102;

Matches 637; Conservative 0; Mismatches 360; Indels 9; Gaps 2;

QY 1 CAATTGATGCTTGGCGACGGATCTCAATCAGCTCTACTCAGCTCGCTCAAGGACTCAA 60

DB 3710 CTAATGACATCTTACCGGGGAATAGCAATTTATTACGCGCTCTCCGGGGCTC-- 3767

QY 61 AAGACCGCTGTGACGACCTACTCTCTACCTATCCAGTACAGCGACTTTGCAAAATGGC 120

DB 3768 -TGATCTCTATCGGTGGTGAGCCACTCCCACTCCACTACCGGACTTTCCGTTGGC 3826

QY 121 AGAAG-----GACCAATTATAGACGAGGAGCACTCAACTCACTGAAGAAGCAAC 174

DB 3827 AAAAGCAGGTGCAACAGGAGACCGCAATCAGCGGCACTCGAATACTGGGTCAAGCAGC 3886

QY 175 TCAAGACTCTTCCCGCAGAAAGATCCGACCGACTTTGCCCGCTTGCCTCTCTCTG 234

DB 3887 TCGCAGACACTCGGCGCGGAATTTCTTAACGACTTCCCGGACCAACATACACTCTCG 3946

QY 235 GAGACGAGGTTCGCTACATGTTTACCATCGAGCGGAGCTTACCAGTCCCTTCGAGCCT 294

DB 3947 GTGAAGCAGGTTCCGTGCCAGTGACGATCGAAGCGCAACTGTATCAAGGCTCCAGAAT 4006

QY 295 TCTGCAACGAACACACACAGCACTCTTTTCGTGCTTCTTCTAGCTGCGTTCGTCGCGTC 354

DB 4007 TCTGTAAGTAGAGCAATGACGCTTTCGCCGTGTTGTTAGGGGCTTCCGCGGACCC 4066

QY 355 ATTATCCTCAGCAGCTGTTGAAGACGCTGTATGTTGTACACCAATTTGCGAATCGCAACC 414

DB 4067 ATTATCCTCAGCAGCGCGGAGAGACTCGATCATCGGCACGCCCATCGGAACCGCAACC 4126

QY 415 GACTGAACTGGAGGATATCATCGCTGCTTTGTCATACGCACTGTATGCGAATCAACA 474

DB 4127 GCCAGGAGCTTGAACATATGATCGCTTCTTCGTTCAACCAACCAATGATGCGAATCA 4186

QY 475 TAGATCATCAGATACCTTTTGGGACTTTTGATCAACCAAGTCAAGCTACGACGACGAG 534

DB 4187 TCGACGGCGAGACACTTTTGAAGCCTGTGCGCAANGTTCGGACCAACGCGGCGG 4246

QY 535 CATTTCGAGAACGAGGATATTCGCTTTGAGCGCGTGTGATCAGCACTACAGCCTCGATCCA 594

DB 4247 CATTTCGAGCAACGAGCTGCTTTTGAAGCGCGTGTGAGCGCACTCTTCCAGCTCGA 4306

QY 595 GAGATCTGTCAAGCAGCACTCTCGCAACTCATTTTTCAGTGCAGTGCAGTGCAGAGGACC 654

DB 4307 GAGACCTATCCGAAACCCACTAGCAGCACTACCTTTCGCTCTTCATCTCAACAGGACC 4366

QY 655 TTGGAAGATTCAAGTTCCAGGGTCTCGAGTCCGCTACCTGTGCTGTAGCAAGCGTACACT 714

DB 4367 TCGCAAGTTTCGAGCTGAGGCTCTCGTAGCGGAACCCGCTCGCAACAGGTATACCA 4426

QY 715 GATTTCGATGAGTTCCTCATCTGTTTCAAGAACCGACAGCCTTAAAGGTAGCTCAACT 774

DB 4427 GGTTCGACGTGGAGTTTCACTGTTCGCAAGCGGGAAGACTAAGCGTAACTGGCAT 4486

QY 775 TTGCGATGAGCTGTGTTCAAAATGGAGACTGTGAAATGTCGTCAGAGTATCTTTGAGA 834

DB 4487 TTGCGCGAGATCTTCAAGCCTGAGACCATAGCATGTAGTGGCCATATTTTCCAAA 4546

QY 835 TTCTGAGAAACGGGCTTCAAAAGTTTCGGGACACAGCTCTCAATACTTCTTTCGATGATG 894

DB 4547 TCCTGGGACAAAGCATTCGCGAGCTCGGACTCCAACTCGCTGTCTCCGCTTACCGATG 4606

QY 895 GCATTGTGACTCTTGAATAATGAGTGTTCACGTCACACATGTGCGATATCCCCGAG 954

DB 4607 GGTTCGCGACCTTCGTGCCATGGCTTTCGATCGAGAAGGCGGAGATACCCCGGG 4666

QY 955 AATCGAGCTTGGCTGATGCTTCCAGACCCCAAGTCTCTGCTTACCC 1000

DB 4667 AGTCGAGCGCTCGAGCTTTCGCGACGAGGTGGCGCTCACCC 4712

RESULT 4

COPEPSYNH

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COPEPSYNH 3155 bp DNA linear PLN 03-FEB-1998

C. oligospermum gene encoding synthetase.

X96559

X96559.1 GI:1770179

peptide synthetase; peptidase synthetase.

Cylindrotrichum oligospermum.

Cylindrotrichum oligospermum

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Sordariales; Chaetosphaeriaceae; mitosporic Chaetosphaeriaceae;

Cylindrotrichum.

1 (bases 1 to 3155)

Bernhard, F.

Identification of genes encoding for peptide synthetases from

Gram-negative bacteria and filamentous fungi

Unpublished

2 (bases 1 to 3155)

Bernhard, F.

Direct Submission

Submitted (12-MAR-1996) F. Bernhard, Freie Universitaet Berlin,

Institute of Crystallography, Takustr. 6, 14195 Berlin, FRG

DAGRFEVSWARHOSGALDAVPHHLKPAEDSRVLKPTDHOGRPLKSLNQPILL
 PAGRRELLIREGLQTLPLPPIPIPSQITLIDRMLNANGKVDRLRLAKRITQKSK
 PVEDIVPRNSREGLVCKGTDLVGEVITIDNFFNLGHSMLAKLALRAGLQNLTR
 ISVRDVPQVADLAIVQIRNAPSAHEPIKPADYTPVQPSFAQGRFLDQNLQVTR
 WYMLPLGIRLHSLRDLALATIASALRQREPLRTTPEEDGVCVQVODHRPKDLRI
 IDLSTOKDAYLAVLKEHQTLPDLATEPCGRVALJRLGEEHILSLVMHIIIDGHS
 VEVLDFPMHRYFSALROODMEQLLPLTYRDFPAWQKTEQVQAHQRLQDITWTH
 LASTPAELLTDUPRSLISGRANELPLTIEGRHDKLRAFCRVHQHTLPIVILALAR
 AHRUTGASDATALGTPIANRNPENLMEFFVNTQCMRIAIEENDNFESLVRVRS
 TATSAFANOPEFESSVSSLPGRSDARNPLVQVILAVHSQDGLKLTLEGRDEAV
 DSAISTDFVEFELFHAHDLRSLGSLVYAKELFKLTIESVVSFLFTRALDQPLTP
 LAVPLTDGVEFELASKGLDPRDTPRDIANIVEQOQHVRAFPDALTAVKDATSILTY
 AQLDQSDRLAIWLSRHMPEFLVGLVAPRSCETIIFAMFGIMKANLAYLPDINSFA
 ARKLSILSAGVGNKLVGLVGTAPQENPEVEAVGLQELLAGTLDKTDGNSARFSA
 TSLAYLFTSGSTKPKGVMSHVSRTLRAPNSVITSLKPOGARVAHLANIAFADSIW
 EIAITLLNGALVLDHYTDLCDRTLKEVERESTVITVLLPALLKQVAEIPEITLH
 LDLLYTGGRVGGHGDAMRARSILVKIFMGSGYPTENTVISTIVEADAMFVNGPIH
 KTVNSGAYDRNOOLVPSGVVDELVTGDLARGVTDPSLANKRFLYITVNGESIR
 AYTGDRVRYRPHDLOITFEFRMDQVVKIRGHRIEPEGEVESALSHNSVQDAVICA
 PADQSGAEMVAFVAARNTDEDTQEEAAVQVQGWETHFETAAYSEVKDIOSEVEN
 DFMGWTSMYDGESEIDTDMHEWLNDRMTILDAREPCHVLEIGTGTGMWFLNAKCPG
 FVKSSEDDMEGQNSLOGSAWAAVMAAQCPSLDAMDVKEIAOEAGYVEVSWA
 ROMSNGALDAIEFHPPEKPGARTLIEPTDYEGRNVTLTNRPLNSIQSRRLGTQI
 REKLTLLPPIPIPSRIMLDQVNNNGKIDRKLVRRAIVAPKPSAATRVAPRNE
 IEALRDEPDELGTVEVLNDFDLGHSMLAKLAAVRSRLDHAISTKDVDPV
 LADLAASIORESAPHEPIPORVYTPGAQSFAGRLWFLDOLNGLATWYMLPLAIR
 GOLVAALSALFALERRHTLTTEESDGVQVIVGEARNSDLRVDVSDGDEY
 LEVLRQETVPLDSEPGMRVCLVKTGEEDHLSIVMHIIIDYDGSVLDLRLGELGF
 YSALRQGLDPLHANPLQTPDFAAWQREAKQVQEHQOLGYWSKLVDSLPAILLT
 DLPRPSILSAGRSVDVTIGSVYALQSFCTRSVTFVLLTVFRIHRLTAVDD
 ATIGTPIANRNPETLVGCFVNTQCMRISIAADDNFEGLVROVNRATAAVANDV
 PPRIVYALVPSGRNTSRNPLVOLMEAVQSDVDVDOVLEGLSVNMPGEASPRFME
 FLHPGDKLTGSLVSSDLEQGTIONFVDFQELRSVLDQPLTPISVLPSNALS
 NLESLLLEMTSDYPRDRTVDFRQAICPDSTAVKDSQSLTAQDESDRVA
 AWHERHMAESLVLSRSCETIIFAYFGIMKANLAYLPDVAADARLAAITLDE
 GERLLGAGVPOPGIOIPRLSTAYIAEALSHATVDTVSIPQPSATSIAYVFTSGS
 TGPKGMIEHRIQVLRDNTVNVFPESSGALPVSHFNLAAMDATWELIYAVLNG
 TVCIDIORDMTALAINSTFRKNRAAFETPAFLKQCLAETPELVANLEIHTAGDR
 LDPGDANLAGTKAGGIFNVLGHENTAVSTFVPVGEESTFVNGVPGISNHAZI
 IDRHKIYPAGVYMLLITGDVGARGYTSALNKDRFVIDINGKSTWSRTSKDARI
 PRDQGLEFFGRMDQWIKVGRIEPEVEVELTLLDHKSALAAVTVVRPENGDEPIA
 FITIDADDVQTHAKYIKHLGLTPAYMIPSHLVIDLQMPVDNGVDRKDLALRAET
 VOKRSTYARVDEVEALDCEYSNLEVEVITDGFDFLGHSHLLATKLAARLSR
 QLNRTSVKDVDFPOIADLADIIRGSHRHDPITPATVYTPVQPSFAQGRFLWFLBQL
 NLGASWYLMPEALRMGRPLQTKALAVANALVHRHEALRTTFEDHDGVGVQVOTPKSS
 QDLRIITLSDAVDDTAYLAALKRQTTAFDLISEPGRVSLRLRGLDDDTILSLVMHII
 ISDGTVDVRLQELGQFYSAALIRGQEPKSLPIQYRDFAVWQRENOIKRQAKQL
 KWSQOLADSTCEFTLDLPRFSLSGEADAPMVIDGTVYQLLDFCRTHOVTSFV
 LLAAFTHRYRTGLTDATVGTPIANRNPPELGLTFGVNTQCMRMAISETFPESL
 VQVRLTTEAFANQDVPPEQIVSTLLPGSRDTSRNLVQVMEFALQSOQDLRIQLE
 MDOALETPLSLRDLVHLEFVQGLSLGLYSLDLEFVETIRGLVDVLEIIRGL
 EOPKRLMAMPITDGLTKLRDQGLTVAPKAPRESSVLDLFRQVAAAPDAIYVWS
 SSTLYADLDGSGNKLAHNMQNMAPETLVAFPRSCETIIFAYFGIMKANLAYLP
 DVNAPARLEALSAPVGHKILVLOAHGPELGLTMADELVDQDEALASSSDHEQI
 HASGPTATLAYMFTSGSTGPKGMIDHRSILRLKNSDVVATLPTVPRMANVSLN
 AFDISVQETITALLNGTGLVCLDLYTLTLDKSLILNVFVEAQVNAAMFTVLLKQCLGN
 MPALISRLSVLNVGDRDLADHADAASGLIQDAVYNAYGPTENGWQSTWYKVDVNEPF
 VNGVPIGRSITNSGAYVMDGNOQLVSPGVMEIVYTGDLGARGYDSDALDRFVHT
 IGBENIKAYTGDVRYRPFKDFEIEFFGMQDOVKIRGRIEPAEVEHALHGLDVLH
 DAAVLRKPANQEPMAIPIITSOEDETIEHESNKOVOGGEHFDVRYADIKDLDT
 TFGHDFLQYKVPAPSAALFVNAEVLKPSLAGKARVLVGTALDGLSKDKNLOPELV
 VINSVAQYFPTSEYLIKVVAVVEVPSVKRFFGDIRSOALNRDLGLAARVAILGNA
 SKEQIREKTAELESEEEELLVDPAFFVSLRSQPLNKHVEVLPKLMKATNELSSTRYA
 AVLHISNEEEQLLQIDIDPTAWDFAATQKDSQGLNRLQQGRDDVMTAVGNIPYSK

Query Match 33.9%; Score 339.4; DB 8; Length 46899;
 Best Local Similarity 60.1%; Pred. No. 3.6e-87;
 Matches 603; Conservative 0; Mismatches 391; Indels 9; Gaps 2;

QY	4	TTGATGCTTTCCGACGGGATCTCAATCAGCTTCTACTCAGCTGCCTCAAGGACTCAAAAG	63
Db	30226	TTGATATCCTCGACAGAGGCTCGTCAATCTACTCAGCGCTTACGTGGC---AGGG	30282
QY	64	ACCGGCTGTGAGCACTCACTCTCTACTATCCAGTACAGGACTTTCGAAAATGGCAGA	123
Db	30283	ATCCGTTATCTCAGGTCGAAGCCCTCCCAATACATAATCTGCTGACTTTTCGGCTTGGCAGA	30342
QY	124	AGGA-----CCAAATTCATAGACAGAGAGCAACTCAACTACTGGAAGCAAGCACTCA	177
Db	30343	AGGAAGTCGCCAAGTTGCCGAGCATGAGAGGAGCTCGGCTACTGGGAGAACCAAGTTAG	30402
QY	178	AAGACTCTTCCCGACAGCAAGATCCGACCGACTTTCGCCCGCTGCACTTCTGTCTGGAG	237
Db	30403	CTGACAGTACTCCCGGTGAGCTTCTGACCGACTTTCGCCCGCCCAAGTCTCTGAGTGGGA	30462
QY	238	ACGAGGTTGCGGTACATGTTTACCATCAGCGGAGCTCTACCACTGCTCCTTCGAGCCCTCT	297
Db	30463	AGGCTGGTGTCTATCCCGGTCACTTTCAGGGGCGGGTCTACGAGAAGCTTCTTGAAGTTCT	30522
QY	298	GCAAGCAACACACACGACCTCTTTCGCTGCTTCTTCTAGCTGGCTTCGCTCCGCTCAT	357
Db	30523	CCAAGGAGCCCAAGTAACTCTGTTCTCGGTGCTATTAAACAGCGCTTCCGGGCCACACT	30582
QY	358	ATCGTCTCACAGCTGTTTGAAGAGCGTGTCTATTGGTACACCAATTCGCAATCGCAACCGAC	417
Db	30583	TTGCTCTACTGTTGACAGAGATGCTACGATCGTACCCCAATTCGCAATCGCAACCGCG	30642
QY	418	CTGAAGTGGAGGATATCATCGGCTGCTTTCCTAATAGCGAGTGTATCGCAATCAACATAG	477
Db	30643	CAGAATCTCGAGCATATCATTTGATTTCTTCGCTCAACACCAATGATCGCTTCTCTCTCG	30702
QY	478	ATCATCAGCATACCTTTGGGACTTTGATCAACCAAGTCAAGGCTACGACGACGACGACAT	537
Db	30703	ATACCGGACGACATTCGAATCCCTAGTCAGCATGTTTCGGTTCGGTGGCTACAGATCGCT	30762
QY	538	TCGAGAACGAGGATATTCGCTTTGAGCGCTTGTATCAGCACTACAGCCTTGATTCAGAG	597
Db	30763	ATTCCAATCAGGATATTCCTTTCGAAGGATCGTTCGGCACCTTCTCCCTGGCTCGAGAG	30822
QY	598	ATCTGTCAAGCACACCTCTCGCAACTCATTTTTCGAGTGCATCACAGAGGACCTTG	657
Db	30823	ATGCCCTCAGAGGCCCACTAATCCAGCTTATGTTTTCGCTTGCATCACAGCCAGATCTCG	30882
QY	658	GAAGATTCAAGTTCCAGGCTCTCGAGTCCCTACTCTGCTAGCAAAAGCGTACACTCGAT	717
Db	30883	GGACATTACTCTCGAAGGACTCGACATGAGCGCTCCCAACAGCGTCCGCAACAGCTT	30942
QY	718	TTGACATGAGTTCCTATCTGTTTCAAGAACCCAGACGCTTAAAGGTAGCGTCAACTTTG	777
Db	30943	TCGACATGAGTTCACCTTGTTCGAGAGCTTAAAGCTGAGTGGTTCATTAATCTCTTTCG	31002
QY	778	CCGATGAGTGTCAAAATGGAGACTTGTGAAATGTCTGTCAGAGTATCTTTGAGATTC	837
Db	31003	CCGATGAGTCTTTTCAGGCTGAAACCAATCAACAGCTCGCTGAGTGTGTTCCAGGAGATAC	31062
QY	838	TGAGAAACGGCTTCAAAAGTTCGCGGACACCACTCTCAATCTTCTTTCGACTGATGGCA	897
Db	31063	TCCGACGCGCTTCGACCAACCCCAAGTCTCCATTTCTACTATGCCCCCTGACTGATGGGT	31122
QY	898	TTGTGACTTTGAAAATTTGGATGTTTCTCAAGCTCAAAACATGTCTGACTATCCCCGAGAAT	957
Db	31123	TGATTGATCTCGAGAACTGGGCTTCTGAAATCCAGAGCAGCAACTTCCCTCCGCGACT	31182
QY	958	CGAGCTTGCTGATGCTTCTCCAGACCCCAAGTCTCTGCTTACCC	1000
Db	31183	ACTCGGTTGTCAGCTCTTTCGAGACAGGCTGCTGCCCAATCC	31225

A40406 3973 bp DNA linear PAT 05-MAR-1997


```

DEFINITION Sequence 33 from Patent WO9425606.
ACCESSION A40406
VERSION A40406.1 GI:2296446
KEYWORDS Tolypocladium inflatum.
SOURCE Tolypocladium inflatum.
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Tolypocladium.
REFERENCE 1 (bases 1 to 3973)
AUTHORS Koehner,H.P., Schneider-Scherzer,E., Schoergendorfer,K. and Weber,G.
TITLE RECOMBINANT ALANINE RACEMASE AND GAPDH FROM TOLYPOCLADIUM
JOURNAL Patent: WO 9425606-A 33 10-NOV-1994; SANDOZ AG (AT)
FEATURES             Location/Qualifiers
     1..3973
     /organism="Tolypocladium inflatum"
     /strain="ATCC 34921"
     /db_xref="taxon:29910"
     join(1953..2034,2111..3180)
     /note="unnamed protein product"
     /codon_start=1
     /protein_id="CAA02484.1"
     /db_xref="GI:2296447"
     translation="MKICPILNSKAGHWGDPDLTGSAPDFRSDAITTPSHGMLAIAAG
ATLNDVYGEDQLTRSFEEHMASICGEALFVSGTMAQIALGALSRLCTGYLADA
TSHVHPFAGLGLSGASIQVPANGHYLTDDVERHAVTNNIERLPTSIISLEN
TAGHSVIPLOELRKAWADEQETAIHIDGARIMHAGVGGSLKEIAGCCDAMTISF
KGLAAPGISALVSRDLIARARLRQSIGGVKAGAIYAAQAMIENTFGPDVDIT
RGIORTHLVARANVMWTSRGULLRPVMTNLVWLDUKSAGLDKSTLDTAAQHRDIR
IRAPRIVLHOISTDAMRRLAEVFEVLEKKASQSLRSRAERAGLERL"
BASE COUNT  917 a 1036 c 1103 g  917 t
ORIGIN
Query Match      16.8%; Score 168.6; DB 6; Length 3973;
Best Local Similarity 56.4%; Pred. No. 1.4e-37;
Matches 337; Conservative 0; Mismatches 254; Indels 6; Gaps 1;

QY      4  TTGATGCTTTCGAGCGGATCTCAATCAGCTCTACTCAGCTCGCGCTCAAGGACTCAAAAG 63
Db      598  TCGACATATTTCAGCAGGAGCTTGGAGAACTCTACACGCGCGCTCGCAGGGGAATCGA 539

QY      64  ACCCGCTGTAGCACTACTCTCTTACCTATCCAGTACAGGCACTTTGCAAAATGGCAGA 123
Db      538  TTTCGGCTGTGCTTGGGTGCAATTCCTCAATCAGCTGACGTGACGACTTGGCAGA 479

QY      124  -----AGGACCAATTATAGCAGGAGAGAGCACTCACTACTCTGGAAGAGCACTCA 177
Db      478  ACCGAGCAGCAGGCTCGCTGAGCAGGAAAGGCGCTCGGATCTGGATCGAGCGCTCG 419

QY      178  AAGACTCTTCCCGACAAAGATCCCGACCGGACTTTGCGCGCGCTCGACTTCTGTCTGGAG 237
Db      418  ATAAACACACACCGCGCGAGCTCTCTACAGAGCTTCCCGCGCGACCTTCCATCTGGCG 359

QY      238  AGCAGGTTGGTACATGTTACCATTCAGGCGGAGCTCTACAGTCCCTTCGAGCTTCT 297
Db      358  AAATCGGGAAGATCTCTTCCAGATCGATGATGATGATGATGATGATGATGATGATGAT 299

QY      298  GCAACGACACACACAGCACTTTTCGCTGCTTCTTCTAGCTCGCTTCTGCTGCTGCTCAT 357
Db      298  GCGGCTCCGACGAATACCGCTTACGCGCTGCTGCTGCGAGCGCTTTCGCGTGGCGACT 239

QY      358  ATCGCTCACAGCTGTGAAGACGCTGTCTTGTGTTACCAATTCGCAATTCGCAATCGCAC 417
Db      238  TTCGCTCAGTGGAGCCGAGGATGCAACCATCGGAGCGCGCTTTCGCAACCGCGACGGC 179

QY      418  CTGAACATGAGATATATCGGCTGCTTTGTCAATACGCACTGTTATGCAATCAACATAG 477
Db      178  CGGAGCTGGAGAAATATGTTGGCTGCTTGGCCACTCTGAGTGCATGCGAGTCTGCTGCT 119

QY      478  ATCATCAGATACCTTTGGGACTTTGATCAACCAAGTCAAGCTACGAGCAGCAGGAT 537
Db      118  ACGAGGACGACACCTTCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 59

```

```

QY      538  TCGAGAACGAGGATATTCGTTTGAGCGCGTGTGTATCAGCACTCAGCGCTGATCCA 594
Db      58  ATGCAACCGGAGCGTCCCTTTGAGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2

RESULT 9
LOCUS      AF204805
DEFINITION Nostoc sp. GSV224 40989 bp DNA linear BCT 15-DEC-1999
complete sequence.
ACCESSION  AF204805
VERSION     AF204805.2 GI:6563396
KEYWORDS
SOURCE      Nostoc sp. GSV224.
ORGANISM    Nostoc sp. GSV224
Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
REFERENCE 1 (bases 1 to 40989)
AUTHORS    Hoffmann,D., Hevel,J.M. and Moore,R.E.
TITLE       Characterization of the nostopeptolide biosynthetic gene cluster of
Nostoc sp. GSV224
JOURNAL     Unpublished
REFERENCE 2 (bases 1 to 16388)
AUTHORS    Hoffmann,D., Hevel,J.M. and Moore,R.E.
TITLE       Direct Submission
JOURNAL     Submitted (13-NOV-1999) Chemistry, University of Hawaii at Manoa,
2545 McCarthy Mall, Honolulu, HI 96822, USA
REFERENCE 3 (bases 1 to 40989)
AUTHORS    Hoffmann,D., Hevel,J.M. and Moore,R.E.
TITLE       Direct Submission
JOURNAL     Submitted (13-DEC-1999) Chemistry, University of Hawaii at Manoa,
2545 McCarthy Mall, Honolulu, HI 96822, USA
REMARK      Sequence update by submitter
FEATURES    On Dec 13, 1999 this sequence version replaced gi:6538756.
             Location/Qualifiers
             1..40989
             /organism="Nostoc sp. GSV224"
             /strain="GSV224"
             /db_xref="taxon:76334"
             1370..14509
             /gene="nosa"
             1370..14509
             /gene="nosa"
             /function="activation of L-isoleucine, L-serine, (2S,
             4S)-4-methylproline and leucine"
             /note="peptide synthetase; contains 4 modules"
             /codon_start=1
             /evidence=experimental
             /transl_table=11
             /product="Nosa"
             /db_xref="GI:6563397"
             /translation="MNVIGKINPLSYGOEAMFVSQIATGIAVNIETAKINSLN
             IAAVRWEKIEHPILRTYTHIEGRPVQVNOOHKFNLDVANASDWSEKOLKEKI
             FAITDRPNEKSDILRNLFTRSKHEHLLTTHHIAAGDMWSFDLLSEFOTLTAKE
             IEQVSEQIOTAPDSLTNSKSYDFRMQSEVLSARGEKLMQYQKLAGELPTLN
             PADKPRPVGTGTGYSGLKDLRIQKLYLALASGTSLYQVLLAAFFVLLRYTNO
             TDILIGSPMRGWGREFKEIVGVFNANLVRSVOENATFEFLAQVSKTVREAQHQ
             DYPFELLAQLOPORDISRSPLCQVFTQAHRCPEPTNSLHSGEOLLEIKPVLGG
             HQGADFDLNLMVEAQVQLQWQNTDLPASTITRMAGHFVTLLEGIYVNPQEQI
             WQLPLTTEQQLQWNTDGVYLNKCTHQLFEQVDSTPNVAVYVNEQQLTYQ
             QLNRYANGLAYLQSLGVENVLGLCVRSLEMYVGLGLKAGGAVPLDPEYPAE
             RLSFMTDTHVKVLLTQKLVESLPHQARVCLDTDWHISIQANODNLNSTVAENL
             SYVITSGTGTGKGVATTHQAVNRLVNTNYIOLTADCEVAOANAIADFAATFRIWG
             ALLSGAKLVITKSVLLPPEFVNIHKHKSIVLFTTALFNQLASVVPQAFSSRLYL
             LFGGADPKWVLELDKGAQPHLLHVGPTENTTFSSVLVLEDTATTATPIGRAI
             ANQIYILDQSLQVPVGPGEIYLGGAQVNOOHKFNLDVANASDWSEKOLKEKI
             EAGGRQKVEGSKLYKTGLARYLADGNIEYLRIDNQVKIRGFIELSEIARLSQH
             GDVQVSCVIVREDTPGDKRLVAVVHONCKPTISELROFLKAKLPDYMVPAIVILE
             SILPTPNKVDORRALPTDLHSEQODKFKVAPNPIEEMALLWTQVLVLELLGTHDNF
             FELGHSLLATQLVSRIRNFKVELPLRFELFAELAEALQIRQLQOQLELSAPRI
             LPKVRDAELPLSYAQORLWFLDQLEPNSAFNIPMALRVLGNLPALQESLLEIHR
             HEALRTNFTVDGKPSQIOTRREGQGTNRREGQTVSIVDFKDLSTNEQEIASKLARQ

```

```

QAIEPFDLALIRTTLLVLSSETEHILLVCMHHVVSQSGMVGVFVQELAAALYDAQO
GQSSSLADPQYADFAIQWREMLQGNVLQNTYWOQOLANAPSLISLPDRPPAY
QTEAGAYKGFALSAELTGLRLTQSKQGVTFLEMLLAARDILLRYTGAEDILVGSPI
ANDRESEGLIGFVNTLVMTDLAGNPSSELLTRREAMEAYSHQHPPEMLVE
ALQPERDLSHTPLFQVMTLNAPISTELNOLSVSSSTIEGTTAKFDLTLSMENTAT
GVVWVEYSTDLFDASTTERMTGHVFTLIESIANPQOSISQPLLTTEVEQHLISLW
NNTQVDPDLCLHQLFEEOVALTPDAVAVFENQOLTYOOLNCRANELAHYLOSIGV
EADVLVGLTCEVRSLLMVVGLIGLKAGGAYVLPDPDPOERLSFMLEDAOLRVLLTQH
QLEKLPQHOQGVCLDMDQFISQSOENLITVQASNLAYITYTSGSTGKPKGAMK
THLGICLMMQOAYQALTALDCILQTPSEFDSVMEFWPLTYITSGSTGKPKGAMK
DSAYLVNILEQQVTHVHFVPSMLQVLEQNLNCSLRKRVCSGEALPVELQEREF
ARLCEHLNLYGTEAAIDYVWQCFTNGHLRTVPIGRAIANQTYILDEHLQPPVVG
VAGELHAGVLAGKYNRPDLTIDKTFPNFSEVGEQSGAKILPNQSOLVNPQO
LYTKGDLARYLPDQGTIEYIGIDNOVKIRGPIREGEIVELLNOCEDQVASCVLARE
TIGKCLVAVVPHQHSKPTNRELQPLKALPYMVPFSAFVILESMPLISNGVDRR
ALPAPDLVSEETKDYAVPPTIEMLAQTSQVLOQLQVGHIDNFFELGGHSLATQL
VSRIRNFVKVLPRLRELFPAATLAELAREIELOQDQLQJASAPPLIPRAENALATQ
AQRLWELDPNPSFYNMISIALRLAGTVEAALEQSFQIIRHREALRNFVTTVDG
QATQIITENWFTSVLIDNLSITTKQETIAEQVLOQOAPPELATQALVRAFVLVL
SETHILLMCTHIVSDGWSIGLFLQELAAALYNAYSOCKSSTLAPLTOYADFAWLR
QWLOGEVLOSQSWOQKASAPTLISLPDRPPSVQYQTHQEBEALSEELTGKLA
RTDLSGNPSSELLGRVREAMEAYSHQHPPEMLVEALQPERDLSHPLFQVNFVLO
NAPTSGLLEGLTVSSPLIKGTTSRFDLTIMQNSPTGLIGWNEYNTDLFDASTIERM
TGFTVLEGLIIANPEERISQPLTLQEQKLIVENWDTQVDYQDQCLHQLPEKOV
DSTDPAITALANANAVFEDQLTYYELNANOLAHYLOSIGVSGDALVGCVERS
LEMVIGLIGLKAGAAVPLDPEYPERLSFILLEDAOVSVLLTQOSILDRLOHQANO
VCLDQALISQSOQDNLISDOVANNLAYITYTSGSTGPKGTAMNOLALNSLILMHR
ENKIPKAGTLQFASINFDVYSQEIPTTCWCGSTLFLIGEEURDUTSALLGLQOKA
IERMFLPVALQAEVAIGELVNSHLREITIAEQLOITPAISOQLSLTDLCTLN
HYGSESHLSETTLTNSVETPLLPVGRPIANAOIYILDRFLQVPPVGPVGELEYIA
GVLSQGYNRQPELTLEKFPNPKRSRAGEQSGRGTENCRLQYKGTGLARVLSG
NIEYLVQDNQVIRGPIREGEIEAELSDLDVQASCAMATPAAGIAREIDPCKRLV
AYIVPOKEQVLTYSFVSELSKLPYEMVPSAIVILDALPLTNGKLDRLRALPDLH
SOLLDKVAPNPLEELISLWAQVLEQVGHIDNFFELGGHSLATQLISRVTSLS
KVPLRSFAATVAELSQOQDLELIALSPILPRAENALPLSAVLRMLFL
DKLNPSAFTPLAURLVGLTARNALTEQSLQEIIRHREALRNFVTDGKPSQIITQ
QINCTLSVDFKLSLQTEQIASQOLVQFDLANEALVRATLVLSSETHALLVCMH
HIVSDGMSGVFQELAAALYNAYSQGEPSPPLTPLOYADFAWNRWQLOVQLOSQ
SYWQHOLANAPLISLPDRPPSMOTFAGAYQOFAKSELTKLTKLOSOQGVTFPM
TLIAAFDILLYRTGEDIIVGSPIANRORSEIEGLIFLNTLVNRLNAGNPSFSE
LGRVREMANDAYSHQNLPEMLVEALQPERDLSHTPLFQVMEVQNPMSQOELLEGT
LTVSPVSTTTKFLDLTGMENTATGLVINEYNTDLFDASTIERMTGHFLLLEGI
IANPEERISQPLTLTEHQHLLVWDTQVDYQDQCLHQLFEEOQVQRTPDAGAVVF
DEPYTERHELMEDQSVLLTQOQLDLRLAHOAKVCLDMDQLLSQSOQDNL
TQOANALAVIYVTSQSGPKVLLIAHOGLLNVFHWQNTFTLDRATQAGTAF
DAVWELPYLTAGASIYLVKSEFSLVKLRDMLISKKITISFLTPPLAQELLSLEW
PTESLANTILTGDKLHQPSDLVPSQLVNNYGPTEVTVTTSGLVVAKEQDQISPS
IGRAIANQTYILDNQLQVPIGIPGELHAGVLAGKYLNRPELTERKPIPNPKRS
REAGEQSGRGAKLTPNSQSLVSPRLYKTDGLARYPDGNIYGLRIDNOVKIRGPI
ELAEIALLQHDQDVOICCVIAREETPGNKRIVAVVVPKEVTPPTDELROFLDKLP
GYMPSVFWMLLESPLTPNGKDRRALPAPDLOELSNVMPNTEVERIIAGTWKTL
EYKVGIYNFTELGHGSLLLVRINQOLQEKGLEVISVDMFNPIPTIHLRSQVLSLTKL
NEDTLIKONRSPQSSIETKAFRNQOQLRQOYLSQRKGRKRL"
14506..18240
/gene="nosB"
/db_xref="GI:6563398"
/gene="nosB"
/notes="polyketide synthase type I"
/codon_start=11
/transl_table=11
/product="NosB"
/protein_id="AAF15892.2"
/db_xref="GI:6563398"
/transl_table="MNSLNTDNNSEITAIISLAGRFPKAKDIDSWFNLVQGVESI
SRLEDELLTSGVSLDLNPNVYKASAVLSDIELFDANFFAYSAKEALIDPQORLF
LELANEIEKAGDDEHTYNGLVYAGVGMNRYLVNLYPHHOLLETVDPLQTLISND
KDFLTPRVAKNLGTAVNVQTAGTSLVAVHLACQSLNLYECMDALVGLTSLTSD
KITYLHOGMILSPDGHCRADFADAKAGGTIASSGATVVLKRLKDAIARDHITHAIK
SAINDGAMKVGFTAPSVSGOAAVIAEQAIAGVDAEITSYIEAHTGTPLGPDIEIA
ALTOAFSOTTDKKGFCAGLSKTLNGLDITAGVAGLTKTVLALONKMLPPSLHFEIP
NPKIDFANSPPVNTLTWRNTTPRRAGVSSFGIGGTGNAHVILEEAPMFEQSGRGA

```

gene

CDS

```

Query Match 10.6%; Score 106.2; DB 1; Length 40989;
Best Local Similarity 52.5%; Pred. No. 2.5e-19;
Matches 309; Conservative 0; Mismatches 268; Indels 12; Gaps 3;
QY 61 AAGACCCGCTGTACGACCTACTCTCTACCTATCATACAGCGACTTGGCAAAATGCC 120
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5154 AAGTGTACCGCGTCATCTTTAGCACCGCTACCAATTCAGATTCGATCTGCG 5213
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 AGAAGGA-----CCAAATTCATAGACGAGAGCAACTCACTACTCGAAGCAAC 174
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

gene

CDS

```

GEOGRNOLLVLSAKTASALEKATANLITHLKEHPDLGDVAVYTLNSGRGFNFRM
LICODLDAVATLSLEAQVFSNYTEITERSVFWFQCGSQGVYNNMAREIVTETVF
KEVDYCEVKLPLGLIDIRHLIYPSSEKIDEASKOQTAQVAPAFIVEALPKLW
QSYGVOQAAGHSIGEYVAATLAEVSELEADLSVAAARGOMMOOLPTGCAMLSILPA
DKIKSLGPELSVAIINQPSQVVCVSTAAIDTLQNLQAAGIECRLQHTSHAFSQM
MEPLIKAEYRVKVTLPKPLPYISNLTGNMTVTQATNPDYIAQHLRSTVLFAQV
EKLALPQVLEVGPNLTTLVAKRHPDKAATQTVLTVSRHPQSKDSQDNHVLFTLGL
QMLAGVNDLWEGFYSHOEHYRPLPTYPFERFYIDPQRTAWGQALQALPTSQLW
TSLVQSQONQANRVEFNELTYOENRWOLDRCLTAYINSAPRGATFNAEEKYSLE
NLLAQCHISPRYQOLFSRWLQILVQGOQREHGLTGLLPCSDYINHELEEVRF
ATSLVLDLIVORGENLAALVVEQPELIEFNELVYPRKNGSGSESPUNTYISLH
SSLGVKRLLEPQASAAASIKSLRQONKQSEISLQRPDIPQPEKLSALHSRPD
LPNSGTAPRNEEDLADIEWELGKQVHIDNFFKLGGDSLAIQVLSRINFTSI
RLSVANLFESEPTIAEIAKLEKIDPDTSFGASEREETII"
18312..28265
/gene="nosC"
/db_xref="GI:6563399"
/gene="nosC"
/notes="activation of L-leucine, glycine and
L-aspartate"
/notes="peptide synthetase; contains 3 modules"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="NosC"
/protein_id="AAF17280.1"
/db_xref="GI:6563399"
/transl_table="MNTIEQLVDSLKQGVLMWEGEQLRVNAPKGVLTLETRDLLTKN
KAELIFLHEKNADTDFFPKAERPNLPSFAORLMLNLEDPNPSNQOAL
KLHGKLVNALEQSLNKLIDRHEALRTNFTINQPVQVITDSLNLVPIVDSELPE
SERELACQOALTEANRPPDLASSPLIRASVVKLEVEHALVTLHHIVDWSIGVL
MRELATISAFCCNNLSQPLPELQYADFAIWQWLOTEVLTQVLTQWYKOLLKNAIT
LLELTPDRPPIQTFRGEIOYGEFSEVSEANFTROEGATLMLFAAFVTLVLYR
YDHPDLFVEELVKVLPQRLDLSHTPLFQVMEVQNPAPISEVLADLITSSIPQASATA
KFDLTLQSMQNTATGLSVMEYNADLFDASTIERMGHFTOLLGIEVANPQOISQPLR
LLEEQQLLKNVDTGVDYQDKCHLLFEEOVELTPDAVAVVIEHOLHYHELCR
ANGLAKLSGLKADVLVGVICVERSELMVWGLLIGLKAGGAYLPLASDPTERLML
LEDAQVSVLTQQLRIDRPEHTARVCLDMDQVLSQSDNLIPETQNTLAYIYV
TSGSTQGPGLVLAHGLLNVFHWQRTITTLDRKATOLAGTAFDAVWELMPLTGA
GASIVLYKPEILLQDMLKSKITISFLTPMAEOLLSLEMPSTLRTMLTFC
DKLHRYPGVLPLFPVNNYGPTEVTVTTSGLVSNGRDNNISPPIGRPTANVEIYL
DSYLOVPVPGVGLHIGGAGLARGYLNRQQLQKFIISNPFRRSGAGQGRSGETI
NSNRLYKTDGLARYLFDGTIEYLGRIQNVKIRGFIELSEIYAVLSQSDVSGEVI
ARETDPGKRLVAVVVAHQKPAISELROILKAKLPDMVPSAFVILKSLPTNGK
VDRRALPDLDITILEKVAPRTIEMLALLMAOVLKLELVGVINDFFELGHSL
LISYAQORLWFLDQEPNSALYNIPALRLVGLTNOVALEQSLHEIIRAEALRTNFV
TVNGKPSQIITQOTNTVSVVGLKHLSTTEOELATQQLVQQAIEFDFLANQVLRAT
LLVLSHTEHILCYMHHIYSDGSMGVFLOELAAALNAYSAQPSLDRATSLPRRG
TPLAPLQYADFALQWQVLOQVLOQQLKDALPALLSTPTDRPRAVQTF
AGAHQEFALSQKLTQGLISQKQVTLFMTLLAAYDTLRYTGTEDILVGSPIANR
DRSEIELGFFVNTLVNRTNLAGNPSFSELLTRVREAMEAYTYQNLPEMLVEALQ
PORDLSHTPLFQVMEVQVGLVGLVTRTPLMPESRTAKFDLNLPMONTADGIV
GWMEYNTDLFDASTIERMTGHFVALLGELIANPEQISQPLLTTEVQHQHLLVWNTD
GVYDQDLGTYQLFEEOVQRTPDVAVIENQOQTYQLNCRANLQYLVLSGVSD
VLVGVICVERSELMVWGLLIGLKAGGAYVLPDPDPOERLSFILDDAALPVLLTQSSL
ESLPEHTAQVCLDSRQIEENRNLPGSKPENLAYIYVTSQSGPKGVQVSHN
CVVNFELASHCPGLTSDTFCVAVTISFDIAALELYPLIVGAKVWVGVREVASDGA
RLLELQHSSTTVMQATPQWLAVASQSTOOLGMKMLCGGALPOLAHOLLEPGA
QVWNLVGPTEITWISSICOLRNESTOLEARSALASCAALRATAPIGRPTANTQFYILD
SHLOPVPVGVAGELHIGGLGLAKGYLNRQQLQKFIISNPFRRSGAGQGRSGETI

```

```
Db 5214 AAAGAGAGTGGTGCAGGAATGACTGCAAAACCAACTAACTTACTGGCAACAAC 5273
QY 175 T---CAAAGACTCTTCCCGAGCAAGATCCCGACGAGCTTTGCGCGCGCTGCACTTCTGT 231
Db 5274 TGGCAAGCGCACCATCTCTATTGTCGTCGCCACAGACCGCGCCAGACTGCTGTGCAGA 5333
QY 232 CTGGAGACGCGAGTTGCGCTAGCATGTTACATCGAGCGGAGCTTACAGTCCCTTCGAG 291
Db 5334 CTTTGTGCTGGTGCATATCAAAAGTTTGCCCTCTCTGCTGAATCACTAGTGGTTGACAC 5393
QY 292 CTTCTGCAACGAACAACAACAGCAGCTCTTTCGCTGCTTCTAGCTCGCTCCGTGCGG 351
Db 5394 AACGTAGTCAAGAAACAGGGTGACTTTGTCATGACACTTTTGGCAGCGTTTGATAGCT 5453
QY 352 CTCATTATCGCTCACAGCTGTTGAAGAGCGTGTGATTGGTACACCAATTCGCAATCGCA 411
Db 5454 TACTTTACCGTTACACAGCTGACAGAGACATTTTGGTGGGTCTCCCATTCGTACCCGG 5513
QY 412 ACCGACCTGTAAGTGGAGGATATCATCGGCTGCTTTGTCGAATGAGTGTATGCGAATCA 471
Db 5514 ATCGCAGTGAATGAAGGGTTAATTGGCTTTTGTCAATACTTTAGTGATGCGTACTG 5573
QY 472 ACATGATCATCAGCATACCTTTGGGACTTTCATCAACCAAGTCAAGCTACGACGACAG 531
Db 5574 ACTTAGCAGCAACCCAGTTTACGAAATATTGACTCGTGTCCGCGAAATAGCAATGG 5633
QY 532 CAGCATTCGAGAACGAGGATATTCGTTTGGAGCGGTTGTATCATGCACTACAGCCTGGAT 591
Db 5634 AGGCATATTCGATCAGCATTTGCGCTTTTGAATGCTAGTGAACCATTCGAGCCAGAA- 5692
QY 592 CCAGAGATCTGTCAAGCAACACTCTCGGCACACTCATTTTTCGAGTGA 640
Db 5693 --CGGATCTTAGGCATACACCGCTATTCCAAAGTATGTTTACACTTCA 5739

RESULT 10
ASP269505 39436 bp DNA linear BCT 03-AUG-2000
LOCUS Anabaena sp. ORF4 (partial), ORF3, ORF2, ORF1, adpA gene, adpB
DEFINITION gene, adpC gene, adpD gene, adpE gene, adpF gene and adpF gene.
ACCESSION AJ269505
VERSION AJ269505.1 GI:9715728
KEYWORDS acyl-carrier-protein; adpA gene; adpB gene; adpC gene; adpD gene;
adpE gene; adpF gene; halogenase; methyltransferase; ORF1; ORF2;
ORF3; ORF4; peptide synthetase.
SOURCE Anabaena sp. 90.
ORGANISM Anabaena sp. 90
Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
REFERENCE 1 (bases 1 to 39436)
AUTHORS Rouhinen, L., Paulin, L., Suomalainen, S., Hyttiainen, H.,
Buikema, W., Haselkorn, R. and Sivonen, K.
TITLE Genes encoding synthetases of cyclic depsipeptides,
anabaenopeptides, in Anabaena strain 90
JOURNAL Mol. Microbiol. 37 (1), 156-167 (2000)
MEDLINE 20392447
PUBMED 10931313
REFERENCE 2 (bases 1 to 39436)
AUTHORS Paulin, L.
TITLE Direct Submission
JOURNAL Submitted (06-APR-2000) Paulin L., Institute of Biotechnology,
University of Helsinki, P.O.Box 56, 00014 University of Helsinki,
FINLAND
FEATURES
source
    Location/Qualifiers
        1..39436
            /organism="Anabaena sp. 90"
            /strain="90"
            /db_xref="taxon:46234"
            /country="Finland"
            complement(<1..1657)
            /note="ORF4"
            /codon_start=1
            /transl_table=11
            /product="putative MutS protein"
CDS
    1..1657
        /protein_id="CAC01603.1"
        /db_xref="GI:9715733"
        /translation="MANHQNKNCFFIIGECTLPIQCAEILLNQHVIVGIISADASII
        NWAEGKNIPYIKPTDHLGEFLSQQPPDYLFISIVNRVYLPQELPROFAINHDAPL
        /protein_id="CAC01599.1"
        /db_xref="GI:9715729"
        /translation="MGEIRVELRKIRSOITOKLONIIOAKSGAIOEQILITORSDFRVI
        PVKAPKDAIPGIVHDSTSGATLVIEPSSVPLGNOLRIFRKQSTAEARLRRLTLE
        KYAATPDLERLLAIVTTDLDAVAKARYSLWGSNPPRFIRHRODEITLRRLLPPL
        VMOQEGNEVPIVDDLLISQIRVVITIGPNTGKPTVTTLTGLAAIMAKVGLFVPA
        REPMPMFSQLADIDGDESQLQSLSFGSHIRISRILEALEEKGVRSQESDFDA
        OSNGVYRRKEEEEEERKEEEKIKLIPSPQSPVPSQSLVLLDEVGAGTD
        PVGSAIAILLKYLANHTQLTMTFHGELKALKYENPQFENASVEFNETTLPSTYR
        LLGIPGRSNALSIALRLGLPEVVAEAKTOVGEATDEVNOVIAGLEAORRHOETKAA
        EAQKILROAERLYKEVSDKAAALEAREKDLRASQELAVQQAITQAKGEIAQVIRLQK
        GTPNAQDAQATASLSQAQTEPAPPPKPRKAGPMKVGDRVRI"
        2320..4077
            /note="ORF3"
            /codon_start=1
            /transl_table=11
            /product="putative arginyl-tRNA synthetase"
            /protein_id="CAC01600.1"
            /db_xref="GI:9715730"
            /translation="MNATQEQKLKLTQALVAAFGTDYNTDPIILVSASNPKFGDYQA
            NVALSKLKGQPRIAGIEIVSKLDVSDICEPPEIAGPIGFINLKQTSYLEAQLNAI
            QIDARLGIPTATNPQKIVDFSSNPIMAKEMHVGHLRSITIGDSIARILEFCGHVLR
            NHVGMGCTQFGMLITYLREVYPDALTTANALDGLVNFYKAKORFEDDPVQETS
            QEVYRLOAGAEEDTLHAKLLCEQSRQEFQIYDNLNVELIERGESFYAPLLPKYVENL
            ATTGLESQGAQKCVFDGFTNREGNPLPIVQKTDGNTYATDLAALGRIQIDKA
            KRIVYTDGSGSNHFAQFPVARKANWIPEDVHVFPGLVGEDGKKFKTRSDTV
            RLRLDDEAINRARADIBNRQEDERTETEFPQNVAVGVISAVKYAKLSQNRSTSS
            VFSYDKMLSGKNTAPYLLAYAVRIQISROGNIQDTNLGINRQILLKEDAEALTAKH
            LLQDDEVIKVEQDILLNRLCDYLYQLSDKFNKFFYENCPLVKESEPARTSRLMLCDLT
            AKTLKGLSLGKIVLERM"
            complement(4770..5894)
            /note="ORF2"
            /codon_start=1
            /transl_table=11
            /product="putative neutral protease"
            /protein_id="CAC01601.1"
            /db_xref="GI:9715731"
            /translation="MSKAKKLMGVQVQLLSPDNETKAVLEYLCQDSGKIYNSGVYFAR
            QTFTEGKLLTNKFDLIYEKSKLSVAQSPVPMQQTLSVAGSGFSFKRLSLFL
            EQSLHFKPAPDYLTKSLFKVAYPTSGGPPTLINGOLRFLSLGTVVRWFQISDFFL
            TMSNIDPKYKKEFTILPKNGAFYLEISYEVRKROKHDLDINOALSTDLCTADNLAA
            CVDLGNLLIDSRSMKAMQNLNKKQVSTRKEGPEAYWDLNDRVTRKRHNHMKGVNK
            AAKLIIDHCLKYGITGLVIGNWEKPSNANRNQRFVQMPGLKDKDLKQLCDLH
            GIRFQTEESYTSKASFLDGSLEPYVGQKPEGWKALLAR"
            complement(6817..8550)
            /note="ORF1"
            /codon_start=1
            /transl_table=11
            /product="putative patatin-like protein"
            /protein_id="CAC01602.1"
            /db_xref="GI:9715732"
            /translation="MYFLPVLTQTDVILNLASVREGNLDVYIDENIDNVKTFGNGE
            LTIGOMHFEAIVFNGGOLTYLDDHETGSLTGCSSLNKATSPITWGATLHNNVYTGQ
            LANISWNYPCQDQIQTRSSPLVNGEPGLVAYWRNLEGEGTTVDQDSGNTHNTLY
            GDSWNLAQVLFAIAGSSSESGHSPSSPAEIAEIACTTEIPIPDNIQEQEQATPSQE
            ESKQTMNTNATPKYILAIADGGIRGIIPAILITETKQKISLEFLDIAGSSGG
            ILALGTPRLDLAGSDSPVPAQYSAEELIYLEYGAEIFYFEWQLGLEDIFV
            QPKYSGEGREEIKQYFGDSPLENNLKEVFYSYDEIQRIPIFFTNKLEKQTKSKFK
            RKLCIGFTLDAALATATPTVFPYRVSSSHNTNGFTLVDSGLVANNPANAILEA
            CISRQENQOALNMEDILVUSLGTSLTSVYDYDVQVKNMGLLQRTKPLNMVLDGSEV
            VAGELELFEATNKGHKYSYRFQFLKSELEDIDNAKPNRQLQTLANILLIQEKNO
            EIDELCSILSS"
            9359..16135
            /gene="adpA"
            9359..16135
            /gene="adpA"
            /codon_start=1
            /transl_table=11
            /product="peptide synthetase"
            /protein_id="CAC01603.1"
            /db_xref="GI:9715733"
            /translation="MANHQNKNCFFIIGECTLPIQCAEILLNQHVIVGIISADASII
            NWAEGKNIPYIKPTDHLGEFLSQQPPDYLFISIVNRVYLPQELPROFAINHDAPL
```

PRVAGVNTSWALNOKETGCTVTHHMAAMVDAGDILKQVILIDIADETALTALNGKCY
 ESALNAPATQVDELSSGTVFATKVNLMNORTYFRFRKRLRAGGIIISWKRCAITVETDALIR
 ALDFGSPNLPGRKKAIDSNTLFIKVEYGLNSLPYGGTITNIEYTYITVETASYD
 IALRVLAINQSALSYLVEKFGQVGYCDLEPNQVRQIKIEKQDSIVKHEAFWVE
 RLGLTLESITTEAKQTSALHLKEPOVARARMPDEAITLWLQRHPQWHRSDFLAAAF
 ITYLARIQSGSCDIDGKLELORLOGLFESLFAVSVYVNIYDQSFPAALKQREF
 TOLPLVRVDRVTRYPSKLSLSHRGSEQFPVVERVEKLEIDYQSPSLGSDLTFLISS
 DGKCCWFYNTDVLNDSIARMOEQFTFLQGLTEPDQCIAYLPSLQOORREILLE
 WNDTQVDDPKDCIHOLESQVGLTDPDAVAVFENQOYLHELNCRAQIAHQLQSLG
 VGTDLVQIERSVERSLWGLGLIKRAGAYVLPDPEYPTERLTFMLANDLAQVSLVLTQ
 QHLVEKLPNMQEVPQCDTDLVICSSQSPITEVQPGNLAVIYVTSSTGTPKGVN
 QHLSNLCNPMWMOATFPLETKOKVQKTPFGDFAWVEFYAPLLAGGOLLIAPKPGH
 TDSAYLLRLTAQOQVITVOLVPSLLOMLLEOGGIECHSKLHVFCGGEVLPVALLBGL
 LSKDVLNHLNGVTECTDATPCNOREIYAQIIVPGRISNTQYIILDNQLOALPV
 GVGPELHISAGARGYLNRPELTQKFIANPESTYPSGLYKTDGLARYLPNGNIEY
 LGRIDNQVIRKFRILAEIAEAVLNQDDVQVACIAREDDPGNKRLVGYIVPSSQMT
 CYSELPQFLKTPDYPVNIFFVMDLSLPLTNGKINRGALPVPDLOGOGTKYVAP
 RTTEETLSLMAQLKLEQVTHDNFTFGHSLATOLVSRITSFKVELPLRSLF
 APTVAQLSPHIORLOQOELTSPPLIPRAENAEPLSYAQOORLWFLDOLQPNSSLY
 NIPIALRLVGLTNQAALQAQSLSEIIRHEALRNFITVDGTPRQVITQATNTVSVVN
 FRHLSTQAIEAQAQAQIAQSFOLDASEVLRLVLVSETEHILNVMCHHIVSDG
 WMSGVFAELAAALYASIGLSPDAPATPLPIQYADFAIWNMLVGLQVDSQ
 LNTWQOQKDPALLVPTDRPRPAQVTFVQYDFALSVELTQGLMOLSQOQDCTLF
 MTLAAFDTLRYGTGTEIDLVCTPIANRDRSEIEGLIGFVNTLVNRSNLSENPSFA
 ELLTRVRSMAISAHONLPLEMLVEALOPERDLSHTPLFOVMALONPXLSEVELSG
 LSISSLPIEGATAKFDLTAMENTGNGLGVWEYNTDLDLFIESTIERTGHFLLTSSY
 VANPLQISQPLUTEVERKOLLIENHTQVNPVDOCIHQLESFQVETPDPAVAVF
 EEOQYHELNCRANGALHLYSLQSGADVLGVCIVERSEWIMTVGLLTIKAGGAYVP
 LDPEYPERLFTLDTQVKVLLTQOOLVNLKPAHTAQAICVLDLTGKIKTQNSNPV
 NTATSNPLAVIYTSSTGQOPVNLVHHVTRLFAATNSWYKFNFSQDWMTHPSVAF
 DSVWEIAGVLLVGVVYPLLTRSPESFYKLLSOEQVITLNOTPSPAFROLIIOAO
 SIGMSNLRLVIFEGEALLESLOQWFERHGQKQLVNMGYITETTVHVYVRPUSL
 DDNRASVIGRIPDLOVYLDQHLQPVPIGVGEMIVGGEGVTRGYLRHDELTQOR
 LFNRPQRSGGRLRYTGDARYLPNGELEYLRIDQVQKIRGFRIEIGEIEGLIAQ
 HPVAVESVVRDEPSNKLAVYVPOVAQAPTAELRSFLKPKKLDYMPIANVAIL
 ESPLTSNKGIDRRSLVPESRAGIEESVAPRTPVEEKLAAQIAWKLVRVEQVIGHN
 FPELGGHSLATOLVRSKIRTSFKVELPLRSLFAAPVAELAPLIQLOQQNIIEIKVP
 ILPRKKNKS"
 16122..31304
 /gene="adpB"
 16122..31304
 /gene="adpB"
 /codon_start=1
 /transl_table=11
 /product="peptide synthetase"
 /protein_id="CAC01604.1"
 /db_xref="GI:9715734"
 /translation="MQNHNEELPSYAQQRLMFLDLOLPNSPLYNIPIALRLVGLTNLR
 AALQSKLEIIDRHEALRNFMTIDGQPAQIIQTEINMTVSVVDLQHLSTLVLSVEV
 TEQASASQLOAKOATQFPDLASESLVRATLVLSFTEHLLVCMHHIVSDGWSGVF
 AHLEALYNAVYTGQRTPLEPLPIQYADFAIWNMLQGLDLOLQSLSYWQQQLKEAP
 LLSLPTDRPRPAQVTFVQYDFALSVELTQKLSQEQEVTLEMTLLAAFDLTLYR
 YTGSDIVVGAPIANNRYGEIEALIGFFVNTLVMTNLEGNRFSSELLTRVRLMADA
 YSHODLPFMEALOLEDLVSYTPLFVQVFGLOQNPNSQVELTGLTVPLVSVETA
 KCDLTAMENTGTLIGVWEYNTDFTGTEIRMTGHFVTLLEGIVANPLERSOLPL
 LTEVEOQQLVWENDVDYPODKCIHQLEFEQCLRTPDAAVAVFENQOYLHELNCR
 ANQALHLYRSHGVADLVGICVERSEIMVGLLAIKAGGAYVLPDPEYPRQRTQM
 LKDSQVLLTQSFLLDQPIAEQNPQYVICLDQESFSELTNDNPNPOSTNNLAVY
 YTSSTGTPKGVNHEQVLAWSLAWCKTFVQSQRLQFQSFSDLSVAEIAITY
 VSGALCYLEITLPLSVQLVDFLADKISHFSFLSPALSALVQATLPDLOVLSVGE
 ASSAELVRGELHGRFRFNAYGPTESTVAATIAICOPNGRKPPIGRPIANQIYILDEY
 LQVLPVGVPEGLHIGGALGARYLNPPELTQKFIENPDNSKFKIPYNAKDQVNSK
 LYKTDGLARYLPDGNIEYLRIDNQVKIRGFRIEIGEIEAVLSQLDGVQVASCIVRED
 ITGDKRLAYIVPQREVPTVTELRQFLAKLPEYMPNPAFVILESLPTTPNGKIDHR
 ALPTDPLHSELKHKVAPRTPVEEILAQIWAQVMPVQVQIEDNFTSGHSLATQLR
 VSHIRNIFOVELPLRELFAPTAELMPLIHLQOQDLELTSPTPLPRAENAEILLISF
 AQNRWLFQDFPNSALYNIPIALRLVGLTNQAALQSLQEIIRHEALRNLITVDG
 OPROIOTATNTLTVAVDLKSLSTAQETAAQOQVQOALQEPDLKQALVRATLVV
 SYTEHLVCMHHIVSDGWSGVFFVQELATLYNAYSIGOPANLAPLIQYADFLWQR
 QMLAGVLDNHWQOQKDPALLSDLPDRPRPAQVTFVQYDFALSVELTQKLSQEQE
 KLSQEQVTFVWMLLAADFLLTRYTEQSDILVGTPIANNRSEIEGLIGFVNTLV
 RTDLSGNSFABELLTVRLMAMDAYAQDLPEMLVLEVLQEPDLSHAPLFOVDFLLQ
 NDPLSTVELTGLTVSPLETESATKFDLTLMQNTGNGLVGVWEYNTDLDLHSTIERM

gene

CDS

Query Match 9.8%; Score 97.6; DB 1; Length 39436;
 Best Local Similarity 51.5%; Pred. No. 7.8e-17;
 Matches 279; Conservative 0; Mismatches 254; Indels 9; Gaps 2;
 QY 58 CAAAGACCCGCTGTGACGACTCCTCTCTACCTATCCAGTACGACGACTTTGCAAAAT 117
 DB 13392 CAGACGCTACGGAATACCTTTAACTCCACTGCGGATTCTAGTACGAGATTTTGC 13451
 QY 118 GCGAAGAGGACCAATTCAT-----AGACGAGGAGAGAACCACTCACTACTGGAAGAGC 171
 DB 13452 GCGAAGAACTGGTAGTCGGAGAGTACTGCAAGGCAACTGAATTACTGCAACAC 13511
 QY 172 AACTCAAGA---CTCTTCCCCAGCAAGATCCGACCGACTTTGGCCGCCCTGCACTTC 228
 DB 13512 AACTAAAGATGCCCCAGCCCTGTAGTGTGCTGCCACAGACCGACCGACCTGCTGTC 13571
 QY 229 TGTCTGAGACGACGAGTGTGCTACATGTTACCATCGACGCGGAGCTTACACGTCCTTC 288
 DB 13572 AACTCTTCTGAGCGGCATATCAAGATTTTCTTCTGTGGAATTAACCAAGGGTTAA 13631
 QY 289 GAGCCTTCTGCAACGACCAACACGACCTCTTTCTGCTGTTCTTCTAGCTGCGCTGCCGT 348
 DB 13632 TGCAACTGAGTCAGCAACAGAGTGTACTCTGTTTATGACGCTGTTGGCAGCGTTGATA 13691
 QY 349 CGCCTCATTTATCTCTACAGCTGTTGAGACGCTGTGTTGATACACCAATTTGCGGAACT 408
 DB 13692 CCTTACTTTTCTCTACAGCAACAGACATTTTGTGAGGACTTCCAATTTGCTAACCC 13751
 QY 409 GCAACGACCTGAACCTGGAGGATATCATCGGCTGCTTTGTGCAATACGCACTGTATCGAA 468
 DB 13752 GCGATCGCAGTGAATAGAGGGTTAATGGCTTTTGTGCAATCTTTTACTGATGCGGA 13811
 QY 469 TCAACATAGATCATCAGTACCTTTGGGACTTTGATCAACCAAGCTCAAGGCTACGACGA 528
 DB 13812 GTAATTTATCGGAAATCCAGTTTTCGCGAATTACTCACTCGTGTTCGGTCTTGGCTT 13871
 QY 529 CAGCACCATTCGAGACGAGGATATTCGGTTTTCGAGCGGTTTGTATCAGCACTACAGCCTG 588
 DB 13872 TATCCGCTTATGCTCATCAGATTTGCCATTTGGAATGCTGTTGGAAGCATTCGAGCCAG 13931
 QY 589 GA 590
 DB 13932 AA 13933

RESULT 11

AF455810

LOCUS

DEFINITION

Xenorhabdus bovienii T228 ABC transport protein xpsd (xpsd) gene,

partial cds; and peptide synthetase operon, partial sequence.

ACCESSION

AF455810

VERSION

AF455810.1

GI:18034620

KEYWORDS

Xenorhabdus bovienii.

AF455810 15582 bp DNA linear BCT 02-JAN-2002


```
/gene="xpsa"
/note="Region: module 1 thiolation domain motif"
4448..14398
/gene="xpsb"
4448..14398
/gene="xpsb"
/codon_start=1
/transl_table=11
/product="peptide synthetase Xpsb"
/protein_id="AAL57600.1"
/db_xref="GI:18034623"
/transl_table="MNDNELISPLAERKRLLEAKAALKTROOKEITHAOPRDCN
VPLSWAQORWELTQLOPAQATAYHMSAGNLQHLNQNALKAALDOIVARHEILRTT
IVDEGQOQIGLSAGSGLSVQDLSPLFSTEQQAEECAEQADLEPDTGQGLPI
RGRLLRAESHVLLLTQHHIISDGWSNIMLQELSTLYAQFQCOQAEPLPALTLQYA
DYALQWQGLQVLEQDQWRSELQAPVLELPDKPRPTQSQVAGSRHELEPLIF
ALSTELKAFORHGTITFTMLLAGWVLLSRISGOHOLVIGSPVANQRGHELEPLIF
FVNTLALRIQDNPSSVELLARKVKNHALGAYAHODLPEQLVEALPPRSLSGHPIF
QVMLALDNTGQOQYFELDGLHLHELPTRODSAYFDLTLTNDTEQSLVGLDEYASDLF
EHASIERVGLTQILSLMADVDSLRVDDLPITSSORTQLLANFNDAITPYKNAI
HQLFEQVETPKIALVWGETQLSYSELNQRANQLAHSIMASGVHPDDRVAICAERS
LDWVIGVGLKAGASYIPLDNPHPTERLAYMLSDSQPVLMTQOHLKARLPVTNIEV
WALDSEHQTCIASQKDNIDASQGLTSLNAYLVLTSGTGLPKGMTEHQNVHL
IHSOFOMSKLTPHPCVLQFASEFQDNSVAETFPFAIGATVVLPRDHIKYVDETFIF
LONGIVTDLPTAFWHLMAQELISAGYSWPEQLRSVAAGEKAERHLTWLSSPT
QKRWLNTGPTETVWATISVIDENLCTYEDIPGRPIANTRIIYILDORGQVPVIG
VNGEIHISGSGVARGYLNRSELTAERFIQDFSDIPGARMYKTGDLGRQPDFTISY
GRNDFQVIRGFRLEIEIAQATLACAGKDAVIVREDNDGKRLWAYLTIQSGAIL
NAASRLQSLVNDAYMLPFAVTLLEAPLNQNGIKIDRPALPAPDRASVREYAEQ
GEIQQOALQWNLGRICRYDSFELGHSLLTVQVARSRLQSLNIERTLODFT
OPVLDAQSLVATOSTOPMLPANREQALPLSWTOORLWFLAQLOPAQATYANMSG
GLHLOQLNQLAKAALDRIVARHEILRTTIVAEGRARQVIGDADCGFSUTVRDLQ
LSSAQATLIEECAFASHPDFAQGLIRAOQLKIGEOQHVLTLQHHISDGWGL
NVMNLSALYAFQGCQADPLPALAIQAYAVMQRWLQGEORLEKQLSYWRNELQD
APTLEPDKVRESVQSYHGQVFTTSLPNSGLRALSORHGATFTVLLAGWGL
LSRLSGQPDVIGPVPANROYSELEPLIGFANTLRLIKLENDPTVSALLARVKAH
LEAYAHODLPEQLVELOPRSLSHSPIQVLMALDNTSSKQSFELAEISLNPALT
RNSAHFDLTALSDNTENSLECEVYSDLSERIERMAGYLNQLNLAAMVADDNLRA
DLPLMHPTQLTLDNFNTAVYQDKLSQLFQEQVETPDALIAIWEADAQSYAE
LNQRANQALHALTAFQGDPRDRAICIERNLNVMVIGMLKAGAGVYNDPDEPAER
LAYTLNSAPKALLTQOHLQALPEKVLQWQDLGHLNSVAQDVPDPROLGF
PHLAYITVYSGTGLPKGMIEHNVNFTYSQCTSELKSTDRVLQFQASVDFDAG
SEIPTPLSLTILRLPAHRIIPDATESHFLQEQALISVIDLPTAFWQHOQMKAGV
GFSSHVSRTVGGKAEHRHFTWQSMPTRHCHWIDTYGPTETVTSATALTLDGPS
YVTELSIGRLINTHYIILDTQGPVIGVTGEIYIGGAGVARGYLNRBELTAERFV
SDPREPHARMYKTGDLGWRPDGTIVILGRNDQFQVIRGFRLEIEQSLAEQALDGF
VSDAVVAVAREEGSKRLVAYLPKSDVTLDAASLRQVSTHIAEYMLPSAFVILDAF
```

Query Match
Best Local Similarity 49.18; Pred. No. 5.1e-14;
Matches 292; Conservative 0; Mismatches 297; Indels 6; Gaps 2;

Qy 4 TTGATGCTTCGCGCGGATCTCAATCAGCTTACTCAGCTGCCTCAAGGACTCAAAAG 63
Db 8232 TTAATGATTGATGAACGAATTTCTGCGCTGTACCAAGCTTCGGTCAGGGCAAGCTG 8291
Qy 64 ACCGCGTGTAGCACTCACTCTACCTATCCAGTACAGGACTTTCGAAAATGGCAGA 123
Db 8292 ATCCATTGCGCTGCATTGACCAATTCATATGCCGACTA---TGCGGTTTGGCAACGGCAAT 8348
Qy 124 AGGACCAATTCATAGACAGGAGGAAGCAACTCAACTACTGG---AAGAACCAACTCAAG 180
Db 8349 GCGTGAAGGGGAAGGCTGAAAACAGTTGAGTTACTGCGGCAATGAGTTACGAATG 8408
Qy 181 ACTCTTCCCAGCAAGATCCGACCGACTTTGCCGCCCTGCCTCTCTGTCTGGAGACG 240
Db 8409 CCCCACCTTATTAGACTGCTACTGATAAAGTCCGCGCTCAGTACAGAGTTATCATG 8468
Qy 241 CAGTTCGCTACATGTTTACCATCAGCGGAGCTCTACAGTCCCTTCGAGCCTTCGCA 300
Db 8469 GTGATCAGGTCACGCTTCACTTATTCGCGCGGAATTAATTCAGGGTTAAGGCAATGAGCC 8528
Qy 301 ACGAACACACACGACGACCTCTTCTCCTGCTTCTTCTAGCTGCGTTCCTCGCGCGCTCATTC 360

8.8%; Score 87.8; DB 1; Length 15582;
49.18; Pred. No. 5.1e-14;
0; Mismatches 297; Indels 6; Gaps 2;

Qy 4 TTGATGCTTCGCGCGGATCTCAATCAGCTTACTCAGCTGCCTCAAGGACTCAAAAG 63
Db 8232 TTAATGATTGATGAACGAATTTCTGCGCTGTACCAAGCTTCGGTCAGGGCAAGCTG 8291
Qy 64 ACCGCGTGTAGCACTCACTCTACCTATCCAGTACAGGACTTTCGAAAATGGCAGA 123
Db 8292 ATCCATTGCGCTGCATTGACCAATTCATATGCCGACTA---TGCGGTTTGGCAACGGCAAT 8348
Qy 124 AGGACCAATTCATAGACAGGAGGAAGCAACTCAACTACTGG---AAGAACCAACTCAAG 180
Db 8349 GCGTGAAGGGGAAGGCTGAAAACAGTTGAGTTACTGCGGCAATGAGTTACGAATG 8408
Qy 181 ACTCTTCCCAGCAAGATCCGACCGACTTTGCCGCCCTGCCTCTCTGTCTGGAGACG 240
Db 8409 CCCCACCTTATTAGACTGCTACTGATAAAGTCCGCGCTCAGTACAGAGTTATCATG 8468
Qy 241 CAGTTCGCTACATGTTTACCATCAGCGGAGCTCTACAGTCCCTTCGAGCCTTCGCA 300
Db 8469 GTGATCAGGTCACGCTTCACTTATTCGCGCGGAATTAATTCAGGGTTAAGGCAATGAGCC 8528
Qy 301 ACGAACACACACGACGACCTCTTCTCCTGCTTCTTCTAGCTGCGTTCCTCGCGCGCTCATTC 360

8529 AGCGCCACGGTGCAGCGTTTGTTCATGACATTACTGGCGGCTGGGGAATTTTGTGTCTAC 8588
361 GTCTCACAGCTGTTGAAGACGCTGTTCATTGTCACACCAATTCGGAATCGCAACCGACCTG 420
8589 GGTAAAGTGGTCAGCCCGATTTAGTATGTTGCTCCGCTCGCCACCGCCAGTACAGCG 8648
421 AACTGGAGGATATCATCGGCTGCTTTGTTCATACGACGCTGATGCGAATCAACATGATAC 480
8649 AGTTGGAGCCCTTAAATCGCTTCTTCGCCAATACATACACTGGCCCTGAGATCAAGTTAGA 8708
481 ATCAGGATACCTTTGGGACCTTTGATCAACCAAGTCAAGGCTACGACGACGACGATTCG 540
8709 ATAACCGCAGCGCTCAGCGCATTTATGGCAAGAGTGAAGGCTCATGCGCTGGAAGCTCAG 8768
541 AGAAGGAGGATATTCCTGCTTTGAGCGCGTGTGTATCAGCACTACAGCTCGGATCCAG 595
8769 CCCATCAGGATCTGCCATTTGAACAATTAGTGAAGTCTTACACCGCCACGCGAG 8823

RESULT 12
PSPBSC
LOCUS PSPBSC 3671 bp DNA linear BCT 03-AUG-1994
DEFINITION Pseudomonas sp. pbsC gene.
ACCESSION X77699
VERSION X77699.1 GI:455513
KEYWORDS pbsC gene; protein C.
SOURCE Pseudomonas sp.
ORGANISM Bacteria; Proteobacteria.
REFERENCE 1 (bases 1 to 3671)
AUTHORS Adams,C., Dowling,D.N., O'Sullivan,D.J. and O'Gara,F.
TITLE Isolation of a gene (pbsC) required for siderophore biosynthesis in fluorescent Pseudomonas sp. strain M114
JOURNAL Mol. Gen. Genet. 243 (5), 515-524 (1994)
MEDLINE 94268496
PubMed 8208243
REFERENCE 2 (bases 1 to 3671)
AUTHORS Adams,C.
TITLE Direct Submission
JOURNAL Submitted (14-FEB-1994) C. Adams, University College Cork, Microbiology Department, Cork, IRELAND
FEATURES
source
1..3671
/organism="Pseudomonas sp."
/strain="M114"
/db_xref="taxon:306"
1089..3546
/gene="pbsC"
1089..1099
/gene="pbsC"
1135..3546
/gene="pbsC"
/codon_start=1
/transl_table=11
/product="biosynthetic protein C"
/protein_id="CAA54778.1"
/db_xref="GI:455514"
/db_xref="SPTREMBL:Q52531"
/translation="MIEELVLYAGFNLGHGADGALPALIOYADVALLMQRORWAGEQE
ROLAYWQEKLGGEOPVLELPMDRSRTIEQSFAGASHNLILDPAISNALKAFARQENV
LFVLLASQALLHRYSGQADIRVGVNPNANRVEIERLIGFFVNTQVLKADIQOMS
FVLLRQVKTAQQAQAHQDLPEQLVEALEPGLSHSPLFQVFNHQAERRASVET
RLNGLSIKPLEWSQTAQFDLTLLNTTEQAHEIYALVATDLDFAATIERLAQHNAL
LQALVAAPTQRIQGLPMDAQOQLLAQWNPVERVEQPLEQICHOATEQAERYPDAL
AVTYDGRLTVAELNANOWAHVLIARGVDPVVRGVAVERSLDMIVAIIAIVKAGG
AYVLPDPCYDDRLSYMIADSGIELLTQGHLLAQLVPVPSGLACIDHQAPEPNETN
PLCLTTPDLSLAYIYVTSGTGPKGALLPUNVRLFSATHEWFDFGQDQSWTLFHSY
AFDFSWIEFGLLYGKLVVPHDLSRPFEDFYLLCDENVTNLTNPSTPFAKLMQV
ASESTRNSRLYVVFGEALVSLRWFEPFGDRAPTLINMYGITTEVHTVYRPLS
MADEQSHSPNIGEPIDVLSWLDGSLNLPVQGCIGELYTAGLAGLARGYLNQAGMTA
TRFVPPDNFQAGERLYRTGDLARLDGVIEYIGRIDHOVKIRGRFELCEIEAOLL
KHADVREAVVLAVDGISGHATGDVDCRERRTGRRAGRTAFBFDQGAFAFNAGLHGS

BASE COUNT	717 a	1128 c	1135 g	691 t	
ORIGIN					
Query Match	8.6%; Score 86.2; DB 1; Length 3671;				
Best Local Similarity	49.7%; Pred. No. 1.3e-13;				
Matches 248; Conservative	0; Mismatches 248; Indels 3; Gaps 1;				
QY	100	ACAGCGACTTTGCAAAATGGCAGAGGACCAATTCATAGACGAGGAGCAACCACTCAACT	159		
Db	1220	ACTAGCGCTTATGCGCGCGGAGTGGATGGAGCGCGTGAACAGGAGCGCAACTGGCCT	1279		
QY	160	ACTGGAAGCAACTCAAAAGACTCTTCCCCAG---CAAAGATCCCGACCGACTTTGCC	216		
Db	1280	ACTGCAAGAGAAACTCGGTGGCGAACAACCGGTCTGGAACCTGCCGATGGATCGCTCG	1339		
QY	217	GCCCTGCACCTTCTGCTGGAGACCGAGGTTGGGTACATGTTACCATCGACGCGGAGCTCT	276		
Db	1340	GCACCATCGAGCAAAAGCTTCGCTGGCGCCAGCCATACCTGATTCTCGACCGCGCTGA	1399		
QY	277	ACCACTCCCTTCGAGCGCTTCTGCAACACACACACGACCTCTTTCGCTGTTCTTCTAG	336		
Db	1400	GCAACGCGCTCAAAAGCTTCGCCGCGCAAGAAAGCTCACCTGTTCTGCTGGTCTGCTG	1459		
QY	337	CTGCGTTCGCTGCCGCTCATATTACGCTCTACAGCTGTTGAAGAGCGTGTCAATGGTACAC	396		
Db	1460	CCTCGTTCAGCGCTTGTGTCATCGCTACAGCGGTACGCCGACATCCCGCTCGCGCTGC	1519		
QY	397	CAATTGCGAATCGCAACCGACCTCAACTGGAGGATATCATCGGCTGCTTTGTCAATAGC	456		
Db	1520	CGAATGCCAACCGTGGTGGGTAGAGATCGACGCGCTGATCGGCTTCTTCGTCAACACTC	1579		
QY	457	AGTGTATGCGAATCAACATAGATCATCAGGATACCTTTGGGACTTTGTCAACCAAGTCA	516		
Db	1580	AGGTGCTCAAGCGCGACATCGATGGCGAGATGTCTTTGTCGATGCTGCTCGCTCAGGTCA	1639		
QY	517	AGGCTACGACGACGACGATTCGAGACGAGGATATTCGTTTTCAGCGGTTGTATGAC	576		
Db	1640	AACAGACCGCGCAAGAGCGCGCCATCAGGACTTGGCGTTCGAGCAACTGGTCCGAAG	1699		
QY	577	CACACGCTGGATCCAG	595		
Db	1700	CGCTGGAACCGCGCGC	1718		
RESULT 13					
AF047828					
LOCUS					
DEFINITION	Pseudomonas syringae pv. syringae syringomycin synthetase (syre)				
ACCESSION	AF047828				
VERSION	AF047828.1				
KEYWORDS	GI:3510628				
SOURCE	Pseudomonas syringae pv. syringae.				
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.				
REFERENCE	1 (bases 1 to 28587)				
AUTHORS	Guenzi,E., Galli,G., Grgurina,I., Gross,D.C. and Grandi,G.				
TITLE	Link between the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases				
JOURNAL	J. Biol. Chem. 273 (49), 32857-32863 (1998)				
MEDLINE	99047670				
PUBMED	9830033				
REFERENCE	2 (bases 1 to 28587)				
AUTHORS	Guenzi,E. and Grandi,G.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-FEB-1998) Mol. Biology, Chiron/Vaccines, Via Florentina, 1, Siena 53100, Italy				
FEATURES	Location/Qualifiers				
source	1..28587				
	/organism="Pseudomonas syringae pv. syringae"				
	/db_xref="taxon:321"				

/note="pathovar: syringae"	
285..28415	
/gene="syre"	
285..28415	
/gene="syre"	
/note="SyE; peptide synthetase"	
/codon_start=1	
/transl_table=11	
/product="syringomycin synthetase"	
/protein_id="AACB0285.1"	
/db_xref="GI:3510629"	
/translation="MGQVLSKAPFKLEVTDPEDVDGTEQIAIHLRESFRQPFESV	
TGOLWESRLVRCGPHWYFNRLHLEWADVIGIGIAHAHVSADYSGLLKGKTPAEQ	
SYLSFEEDRAYMSSRYKRDRLFWEQSYAQLPVLILRRADFKAGLTNVPASDQVQ	
AMLPALYNALTQFASecnLSVAHVLGVISTYFCRTYGVDEIVVGMVPHNTNARQ	
ATIGFSSVPIRVGDPFHETLVDLMNNTIGTQRHPTIRHOEPIAELNRSLSLAONGR	
HQLDFSLSEFSDGVDVTGSEDPAHVLMIDNGYEQTPMIAFVRDHPHSDHLEHDFN	
NTAYFTIEAQQRQIRIFAMLEAVLNQHDPTLAHFFLMSTSEHQALATFNDAHYPR	
DVLIHQIEQAQAQRDACAAGSDGSLTTPYAEELNQANQLAHRIELFELGVEDPRTVA	
SLRGAASVALLGILKAGGAYPIDPDLPSARQAYMLEDSSPOAVLTTRDLSNPLA	
SDLPVLVDGHNDRALQROQSNPAKALGLQPNHLAVLYLTSGSTGTPKGMNEHL	
GVNRLMADAYQVANSQDRLOKTPCGPDLVWEEFLPLLTGAELVMAPEGGHODPD	
YLAQVMSDAGITLLHFVPSMLDVFLEHSTRDPFQPLRRVLCGSEALPLRAQRRFQQL	
KGVLEHNYGTEAAIDVTAWECRPTDPSVPIRQPIRPIANIQHVLDALQUGPMGVA	
GELHIGIGVARGYLNQPSLAERFTADPSNDPQARLYKTDGVRGMLNAGALEYIGR	
NDFQKILGRILETIGEIEAALKHPVAHVATAREDIQDKRLVAYTYQSAETHAVD	
LEARLSHQOVLPEYMWPAIYVLLAEMLTSGKLDKALPAPDGLASIRGYEAPQG	
EVEFTLASIWADVLKPVQVGRHDHFELEGHSLLAVKLTERMQVGLSADHVILFCOP	
TVAALAAAGGSGSEVVVPAANIPEHCDRTIPQMLPLINLSQAQIDHVVKDMPGGVANV	
QDITYPLAPLAGILYHHISAEQHPYTLKALFALSDRAQDDPSGALQGVINRHDLIR	
TATLWEDLDEPVQVYLRQAEHLVTELLDPVDQVDEQLHQRPDRHRYRLDVTAPL	
RVFSDHPVNDRLMALLCHLVSDATLSVILQEIQAHLGCGNALGSEAVPYRNYVA	
QARLGVEAGAEHAFRDMLGDIDEPTLPFGLDQVQDRGNLBEASVTLAAELNRLRA	
QARQAGYSVASLMHLAWRVLGNVSACEVVFVTLVLRMGAGDGADRALGMFINTP	
LRVDIGATTVAEGLKATHERTALTALGHEHAPVLVLAQRCGVAAPPTFSALLNLSH	
YSVDYRQKPLGCGSSLRHRPQIOPYDLNVDDLGDATFTLTIONVOQISATRTGEYMQV	
ALRNVDALDEHTPQAALNSUSLIPDDERELLTGFDNTHAPYPRDVLTHHQLIEQQAQ	
RPDCAVDSGTLTYAEELNQANQALHRLIELGVEPDTTRVAVSLRRGAEVMVALLG	
ILKAGGAYPIDPDLPAQNYMLSDSAPRAVLTSHELLADPLGVPVALLVARDGDS	
ALLKQGTGPNDKALDLOQNYHLAVLYTSGSTGTPKGMNEHLGVNRLMADAYQ	
VNSODRYLQKTPFGFDVSWEEFLPLLTGAELVMARPSGHQDDPYLAQVISDAGITLL	
HEVPSMLDVFLEHSTRDPFQPLRRVLCGSEALPLALORFEOHLKGVLEHNYLGTEA	
AIDVTAWECRPTDPSVPIRPIANIQHVLDALQQLQPMGVAGELHIGIGVARGY	
LNQPLSAERFIADPFSNDPQARLYKTDGVRGMLNAGALEYIGRNDQFVKVIRGLRIE	
GETIEAALAKHPVAHVATAREDIQDKRLVAYTYTAEHTAVDTELRHQLQOVLPE	
YMPVATYVLLAEMLTSGKLDKALPAPDGDALIIRGYEAPQGETEEQIAYIWQDL	
GVEQVGRHDHFELEGHSLLAVSLIDRLKQDLNNVNTVTPAPSREMAAISQEKQ	
ALFPANRPIHACTOLTDPMLPELSVAQIELIASVPGAAGANTQYIPLAQLEG	
ILFHVILNLERDAYLVBSTIEFDSRLADFLLEGQTVIDRIDVLRSSVHWVGLQVAV	
QVVIHQASPIHTLTLPDEDAISQDRSDPGRLLDLRQAPLLLAYLARPDPSERW	
LLALIDHIMISDHVTELEELIIRLMRGSALPLPPQTRFEVATQTLASPSAHEAY	
FTGLADVDSPTAFPELLEVGQDNDVESELSALSDLCARINTQARSGMSPAVLFW	
VAAQVILARCTGRDDVVGFTAVTGRLQGTGLPFAERAMGMFMTLPVRVQIATQSVQEL	
MATHRDLSELLSHEQASLALACRCSVATGVPLFSSLLNRYHQNESQQLQWGLRLLD	
SAERTNYPCLSYNDYSGDLGLIHSVPQDPORLCAWMOCALETDALATPOKEV	
TOLDVLPAAERNLLTFENOTRODYPDCLCIQLHFEAQVTRQTPDATALVAQAQRLSYA	
ELNRQANRLAHLHILGIGIPDDRVAICVERGEMVMVGLLGVLKAGAAVYPLDPAYPE	
KLATNIEDSQPALMTORHLQETPLTLPVLDDDDQRTTFERDDNPPVRCALGVNR	
LAVVITYTSGTGNPKGMVIEHRLVNSVDAARLFDLSPTDVTVOQNTLNFDSVEEI	
FPALLAGATILPSREIFGSEGTENHINTVTLHLTAHHTVLVAEHMKQPOVAERQLQ	
HVRLINTYGDALSQAQLKLDWEVPARLTRINTYGTTEATVTSCTAYVSHDAAAGSEG	
SGNATIGKPWANTRYLLDHAQOQPVPGVAGEIYIGGQVARGYLNLEWNAERFLAD	
PFSPSPDARMYKTGDLARYNADGRIEYLGNDPQVVKRGRIELGRIEARLNCNCTGVK	
EAVYIAREDNPGDKRLVAYVVAQPSQLTRAADRLAELAPQLAEYMLPSAFVLLDELPL	
TPNRLDKKALPAPDADALISRGTEAPQGETETQVIAIWDDLLIGISQVGRHDFELG	
GHSLLVASLIERLKQGLNINVKTVFTAPTREMAIAISQDKHVLQVGPANRPAHCT	
QLTPDMLPLVELSVAQTELITSAVGAANNIQIYPLAPLQDGLIFHYLLNREDAYLT	
MRSMIEFDSRLADFLAQTELQVTRDHDILSRSSVHVMGLQVAVVHRRQALPVHTLT	
LTPEDALSOLDRLSOPGRLLDLRQAPLLLAYIARDPNSERWLLALIDHIMISDHTV	
VELLIEELRLMRGQSNDDLPPOPYRDFVQATILASPSAHEAYFTRRLADVDSPTAF	
ELLEVGQDNDVEEAKLNSDLCIRVTOARERGMSPAVLFWHVAQVMACTGRD	
VWFGTVVTYGRLOQTAGARGMGMFMTLPVRVOLTQGAOELVMDLSLSLSHGEQ	
ASLALAQRCSSVATGVPLFSSLLNRYHQGDNRLQWPMGRLLDGTERTNYPLCLSVND	

/note="pathovar: syringae"
285..28415
/gene="syre"
285..28415
/gene="syre"
/note="Syre; peptide synthetase"
/codon_start=1
/transl_table=11
/product="syringomycin synthetase"
/protein_id="AAC80285.1"
/db_xref="GI:3510629"
/translation="MGQRLVSKAPFKLEVTDFPEVDGTEQAIAHLRESFPQFESV
TGLWESLVRGCPQLHFWNRLHLLVADGICAVLIAHAYSDAYSGLGKGTLEAGP
SYLSLEEDRAYMASRYKRDRLFWESYQAPPPVLLQRRADFVAGLTNVPAPSQVQ
AMLPALYNALQFASACNLSVAHLVIGVISTYFCFTGVGDEIVGMPVHNHTNAQK
ATIGFSSVSPIRVGDFPHETILMNIGTQLRRTYRHFPIAEILNLSLAQNGR
HOLFVLSFSFSDVTFGSDPARVLMONGYEQTMAIEFVRYHPEDTHLDFN
NTAYFTFEAQRVQORIPAMLEAVLNQDTPLAHFLPMSTSEHQALATFNDAHPYPR
DYLIIHQLEQAAQRPDACA VRGDSGPLTYAELNQANQLAHLRIELGVEPTRVAV
SURGAEMVALLGLKAGGATVPIDPLPSARQATMLEDSQVAVLTITRDUSDNLP
SDPLVLDGHNDRQAQRQSNFPAKALQPNHLAYLYTSGTGTGPKGMNEHL
GVNRLWARDAYQVNSQDRVLQKTCGFDLSWEEFFLLPTGAELMAPPGHGDPD
YLAQVNSDAGITLLHFVPSMLDVFLEHRSRDFPQLRRVLCGSEALPRALQRRFEQOL
KGVHNLGPTTEAALDVTAWECRPTDGDSPVIGRPIANIOMHVLDAQLOQPMGVA
CELHTGGVARGYLNPQLSAERLADPFSNDQARLYKTGDVGRWLANGALEVLR
NDFQVKGIRLEIGIEIAKLAHFAVHEAVTAREDIPGKRLVAYTQSEHTAVD
LEALSHLQQLPEYVPAIYVLLAMPLTNSGKLDKALPAPDGDALISYEAPQG
EYETLLASIWADVKEQVRHDFEELGSHSLAVKJETERMROVGLSADHVLFQGP
TVAAALAAAGGSEVVVPAIPEHCDRITPQMLPLINLSQAQIDHVKMDPGGVANV
ODIYPLAPLAGILYHHSAEQDPVTLKALFALSDRAQLDDFSGALQGVINRHDILR
TAVLMEDDEPVQVVLROAELHVTLLDPVDPVDEQLHQRDRHHYBLDVRTAPLM
RIVFSDPVNDRWLMLCHLHLVSDATSLVSLTQELQIHLQCGNALGAVPIRNYVA
QARQGVSAQSEAFRDMGLDIDEPTLPGLQVQDQDRGNLEEA SVTLAAELNRLRA
QARQGVSAVSLHMLAWRLGNVSAQVFGVTLGRMQAGDAGRALGMFINRLP
LRVDIGATTVAEGLKATHERLTALLGHEHAPVLAQRCGVAAPPLFSALLNRYHSV
VSDVQKPLGGLSGSLRHPYQVLDLNVDDLGAFTLTIOAQVQISATRIGETYMOV
ALRNLVDALEHTPOALNSLSILPDERELLITGFNDTAHPYPRDVLHOLLEQAAQ
RPDACA VRGDSGPLTYAELNQANQLAHLRIELGVEPTRVAVSLRGAEMVALLG
ILKAGGATVPIDPLPSARQATMLEDSQVAVLTITRDUSDNLP
ALLKQPTGNPKADLQPNHLAYLYTSGTGTGPKGMNEHLGVNRLWARDAYQ
VNSQDRVLQKTCGFDLSWEEFFLLPTGAELMAPRSPGHQDPDLAQVLSIDAGITLL
HEVPSMLDVFLEHRSRDFPQLRRVLCGSEALPRALQRRFEQHLGVEHNLGPTTEA
AIDVTAWECRPTDGDSPVIGRPIANIOMHVLDAQLOQPMGVAELHIGGIVARGY
LNQPLSARFIADPFSNDQARLYKTGDVGRWLANGALEYLRNDQFQVKIRGLRIE
GEIEAALAKHAPVHEAVTAREDIPGKRLVAYTQSEHTAVDLETKHQVQVLP
YMPAIYVLLAMPLTNSGKLDKALPAPDGDALISRGYEAQPEEQIAVWQDILL
GVQVGRHDHFFELGGLSLLAVSLRQKQNLNNTVFTAPSVREMAAISQEKQ
ALFRAPANRIPAHCTQLTDPMLPLVLSVAQIELTASAVPGGAANTQIYPLAQEG
ILFHYLLNLRDAYLVRSTIEFDSRARLDALFEGLOTVIDRHDVLRSSVHWGLQVAV
QVYHQAQSPHITLIRPDEDALSOLDRLSDPGRRLDLRQAPLLLAYLARPDSERW
LIALIDHMHSDHVTLELLEELRLMRQSAELLPPQPYREFVAQTLPASPSAHEAY
FTGRLADVDSPTAPFELLEVQGDNDVESELASLSDLCARITQARERGMSPAVLFH
VAAQVLA RCTGRDDVFGTAVTGRLOQTGAEARGMFMNTLPVRVQATLSQVQELV
MAERTHLSHELLSHEQASALAAQRCSSVATGVPFLSLLNRYHQNSDQLQWGLRLD
SABRTNYPCLSVNDYSDGLLIIHSVPADPQRCLAMMOCALEOLDALHTPOKEV
TOLDVLPAAERNLLETENOTQDPTDLCIOHLEFAOVRTOPDAIYAVFAQRLSYA
ELNRANRLAHLHIGLIGCPDDRVAICTVERGVEMVGLGLVLCAGAAVPLDPAIPAE
RLAYMIEDSQPALMTQRHLLEPTLTPVLVLLDDQKRTETERDNDNPFVVALGVN
LAYIYVSGTGNPKGVMEHRLGVNYSVDAARLFDLSPTDTVQTNLFNDSLVEEI
FPALLAGATLTPSREIFGSEGTENHINPTVHLTAHMHVTLVAEHWKQVPAEORLQ
HRLINVTGDALSQKLLWDEVRPAHTRLINTYGTPTATVSTCTYAAHAAAGSEAG
SGNATIGKPMANTRYLLDHAHQVPYVAGVEIYIGGQGVARGYLNLEVRNARLEAD
PPSESPDARMYKTGLARYMADGRIEYLRNDQFQVKIRGFRIELEIARLNCITGVK
EAVVITAREDNPGDKRLVAYVVAQPSQTAADRLAEALPQALAEIPLPSAFVLLDEPL
TPNRKLDKALPAPDADALISRGYEAQPEEQETQIYVIMQDLLEQVRHDFHFFEL
GHSLLASLIERLQKQNLNNTVFTAPVREMAAISQKHKHVLFPVPAHRIPAHCT
OLTDPMLPLVLSVAQIELTISAVAGGAANIODIYPLAPLQDGLFHYLNLRDAYL
MRMIEFDSRARLDALFEGLOTVIDRHDILRSSVHWGLPQAVQVYHQAQVPLHTL
LTPEDALSOLDRLSDPGRRLDLRQAPLLLAYLARPDSERWLALIDHMHSDHVT
VELILEIRLLMRGOSADLLPPQPYREFVAQTLPASPSAHEAYFTRRLADVSPFAP
ELVEYQGDNDVEEAKLAINSDMCIRITQARERGMSPAVLFHFAVQAQVMACTGRDD
VYFGTVTGRLOQTAGARMGFMNTLPVRVOLTQGAQELVMAHVAQMSLLESHQ
ASLALAQRCSSVATGVPFLSSLLNRYHQEDNRLQWPMGRLLDGTERTNYPLCLSVND

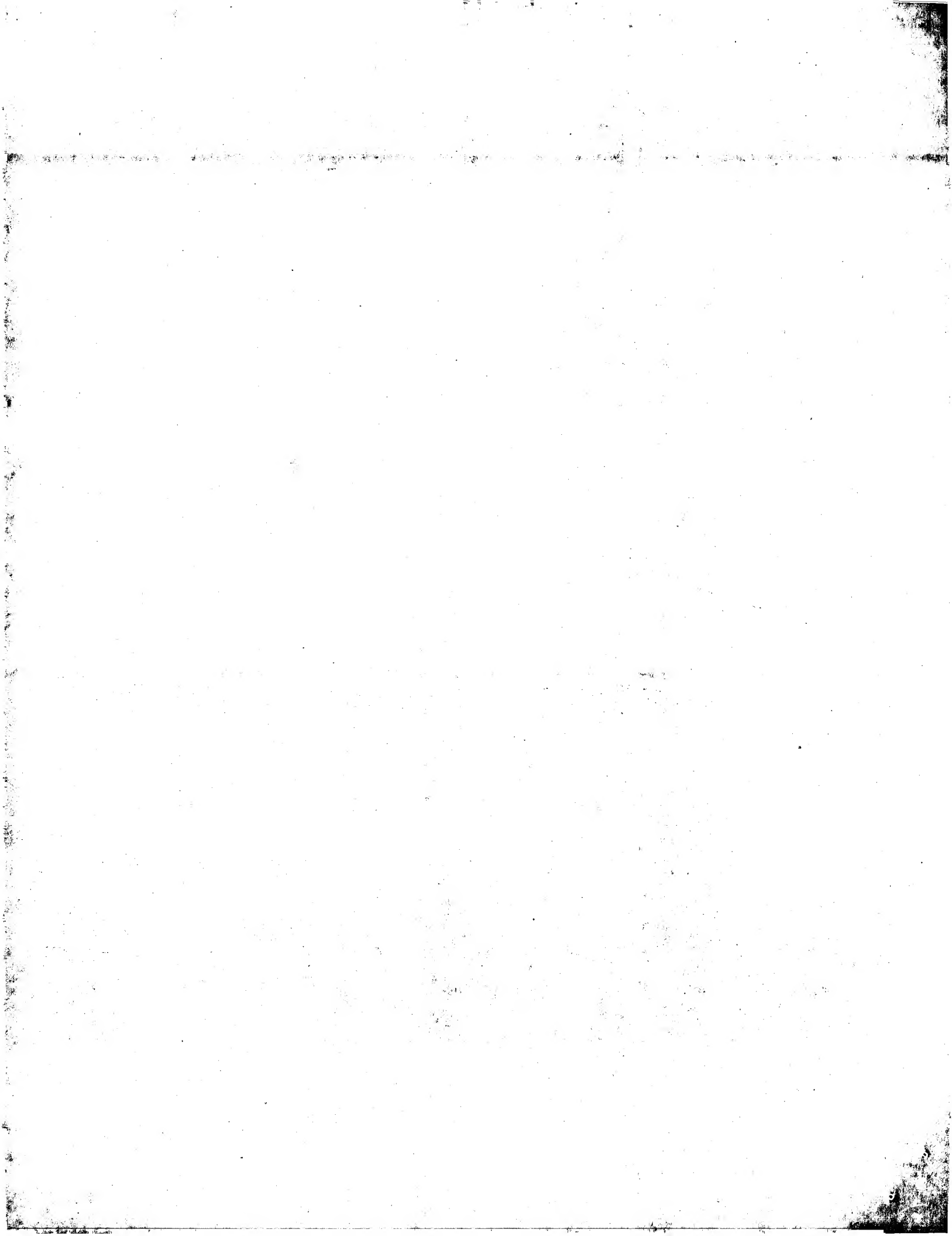
REFERENCE 1 (bases 1 to 37856)
AUTHORS Beyer, S. and Mueller, R. J.
JOURNAL Patent: DE 19846493-A 13-APR-2000;
BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES
source 1..37856
/organism="Polyangium cellululosum"
/db_xref="taxon:56"
BASE COUNT 5655 a 13666 c 12913 g 5622 t
ORIGIN

Query Match 8.6%; Score 85.6; DB 1; Length 37856;
Best Local Similarity 50.2%; Pred. No. 2.4e-13;
Matches 270; Conservative 0; Mismatches 259; Indels 9; Gaps 2;
Qy 61 AAGACCCGCTGTACGACTACTCTCTTACCTATCCAGTACAGGACTTTTCAAAATGGC 120
Db 20535 AGGCGCGCGCTCCCGCTCCCGGAGCTGCGGATCAATAGCCGACTTCGCGGTGGC 20594
Qy 121 AGAAGGA-----CCAATTCAATAGAGGAGAGCAACTCACTACTTGGAGAGCAAC 174
Db 20595 AGCGGGAGCTCTCCAGGGCGAAGTCTGGAATCGCACCTCGGGTACTTGGAGAGCACC 20654
Qy 175 TCAAGACTCTTCC---CCAGCAAGATCCCGACCGACTTTTCCCGCCCTGCACCTTCTGT 231
Db 20655 TCCGCGCGCGCTCCCGCTCCCGGAGCTTCCGATGGACCGCGCCCGCGCGCAGA 20714
Qy 232 CTGGAGAGCGAGGTTCGATACATGTTACATCGACGCGAGCTCTACCAATTCCTTCGAG 291
Db 20715 CGTTCCGGGGCTCCCGCGCGGTTCGACTCCCACTCTCCCTGCAACAGCGGTGAGG 20774
Qy 292 CCTTCTCAAGAACACACAGACCTCTTTTCGCTGTTCTTCTAGCTGCGTTCCGTGCCG 351
Db 20775 CGTTCAGCGGAGAGGAGCGGACCCCTTTCATGACGCTGCTCAGCGGTTCAGCTGC 20834
Qy 352 CTCATTATCTCTCACAGCTGTTGAAGACGCTGTTGATGGTACACCAATTCGGAATCGCA 411
Db 20835 TGCTCTCGGTTATCGCGCGGAGAGCGATCTGGTGGTGGACCGCCATCCGGAATCGCA 20894
Qy 412 ACCGACTGAACAGGAGATATATCGGCTGCTTTGTCATACGACGAGTGTATGCAATCA 471
Db 20895 CCCGAGCAGAGCTGGAGGGGCTGATCGGCTTCTTCGTCACATGCTGGCGCTGCGCATCG 20954
Qy 472 ACATAGATCATCAGATACCTTTGGACTTTTGTATCAACCAAGTCAAGCTACGACGACAG 531
Db 20955 ACCTCGGGGGGAGCCCGAGCTTCCGAGCTGCTCGGGCGGGTGGGAGGTGACGTTGG 21014
Qy 532 CAGCATTTCGAGAACGAGGATATTCGGTTTGGCGCGTTTGTATCAGCACTACAGCCTGG 589
Db 21015 GCGCTACGCGCACCGAGGACCTGCCCTTCGAACGCGTGGTGGAGGAGCTGTACACAGG 21072

RESULT 15
AX024212
LOCUS AX024212 37856 bp DNA linear PAT 15-SEP-2000
DEFINITION Sequence 1 from Patent DE19846493.
ACCESSION AX024212
VERSION AX024212.1 GI:10184520
KEYWORDS
SOURCE Polyangium cellululosum.
ORGANISM Polyangium cellululosum.
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Mycococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 37856)
AUTHORS Beyer, S. and Mueller, R. J.
JOURNAL Patent: DE 19846493-A 13-APR-2000;
BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES
source 1..37856
/organism="Polyangium cellululosum"
/db_xref="taxon:56"
BASE COUNT 5655 a 13666 c 12913 g 5622 t
ORIGIN

Query Match 8.6%; Score 85.6; DB 6; Length 37856;
Best Local Similarity 50.2%; Pred. No. 2.4e-13;
Matches 270; Conservative 0; Mismatches 259; Indels 9; Gaps 2;
Qy 61 AAGACCCGCTGTACGACTACTCTCTTACCTATCCAGTACAGGACTTTTCAAAATGGC 120
Db 20535 AGGCGCGCGCTCCCGCTCCCGGAGCTGCGGATCAATAGCCGACTTCGCGGTGGC 20594
Qy 121 AGAAGGA-----CCAATTCAATAGAGGAGAGCAACTCACTACTTGGAGAGCAAC 174
Db 20595 AGCGGGAGCTCTCCAGGGCGAAGTCTGGAATCGCACCTCGGGTACTTGGAGAGCACC 20654
Qy 175 TCAAGACTCTTCC---CCAGCAAGATCCCGACCGACTTTTCCCGCCCTGCACCTTCTGT 231
Db 20655 TCCGCGCGCGCTCCCGCTCCCGGAGCTTCCGATGGACCGCGCCCGCGCGCAGA 20714
Qy 232 CTGGAGAGCGAGGTTCGATACATGTTACATCGACGCGAGCTCTTACCAATTCCTTCGAG 291
Db 20715 CGTTCCGGGGCTCCCGCTCCCGGAGCTTCCGACTCCCACTCTCCCTGCAACAGCGGTGCAAG 20774
Qy 292 CCTTCTCAAGAACACACAGACCTCTTTTCGCTGTTCTTCTAGCTGCGTTCCGTGCCG 351
Db 20775 CGTTCAGCGGAGAGGAGCGGACCCCTTTCATGACGCTGCTCAGCGGTTCAGCTGC 20834
Qy 352 CTCATTATCTCTCACAGCTGTTGAAGACGCTGTTGATGGTACACCAATTCGGAATCGCA 411
Db 20835 TGCTCTCGGTTATCGCGCGGAGAGCGATCTGGTGGTGGACCGCCATCCGGAATCGCA 20894
Qy 412 ACCGACTGAACAGGAGATATATCGGCTGCTTTGTCATACGACGAGTGTATGCAATCA 471
Db 20895 CCCGAGCAGAGCTGGAGGGGCTGATCGGCTTCTTCGTCACATGCTGGCGCTGCGCATCG 20954
Qy 472 ACATAGATCATCAGATACCTTTGGACTTTTGTATCAACCAAGTCAAGCTACGACGACAG 531
Db 20955 ACCTCGGGGGGAGCCCGAGCTTCCGAGCTGCTCGGGCGGGTGGGAGGTGACGTTGG 21014
Qy 532 CAGCATTTCGAGAACGAGGATATTCGGTTTGGCGCGTTTGTATCAGCACTACAGCCTGG 589
Db 21015 GCGCTACGCGCACCGAGGACCTGCCCTTCGAACGCGTGGTGGAGGAGCTGTACACAGG 21072

Search completed: June 4, 2003, 20:51:10
Job time : 2769 secs



Result	Query No.	Score	Match	Length	DB	ID	Description
	1	1001	100.0	11212	21	AA58762	DNA encoding a cyc
	2	392	39.2	9633	22	AAF79702	Mycella sterilia c
	3	339.4	33.9	46899	15	AA054386	T. niveum Cyclospo
C	4	168.6	16.8	3973	15	AAQ78276	T. niveum alanine-
C	5	168.6	16.8	3973	15	AAQ78281	T. niveum alanine-
	6	85.6	8.6	37856	21	AAAI1992	S. cellulosus DNA
	7	75.6	7.6	13029	23	AA351470	Pseudomonas aerugi
	8	75.4	7.5	7335	22	AAF81367	Quorum sensing con
	9	74.2	7.4	18660	21	AA58472	Nucleotide sequenc

QY		61	AAGACCGCTGTGAGCAGTCACTTCCTCTACTATCCAGTACAGGCAGTTTGCAAATGGC	120
Db		3768	-TGATCCTCATCGGTGGTAGGCCACTCCCACTCCAATCCCGCACTTTTCGGTTGGC	3826
QY		121	AGAAG-----GACCAAITTCATAGACGAGAGAAGCAACTCAACTACTGTGAAGAAGCAAC	174
Db		3827	AAAAGCAGTCTGACAGGAGACCGAACAATGAGCGGCNACTCGAATACTGGTCAAGCAGC	3886
QY		175	TCAAAGACTCTTCCCAGCAAAAGATCCGACCGAGCTTTGCCGCCCTTGCACTTCTGTCTG	234
Db		3887	TCGACAGACAGCTCGGCGCGCGAATTCTTAACCGACATTCCCCCGACCCAACATACTGTCCG	3946
QY		235	GAGACGGAGTTCGGTTCATGTTAACATCGACGGCGAGCTCTACAGTCCCTTCGAGCCT	294
Db		3947	GTGAAGCAGGTTCCGTGCCAGTACGACATCGAAGCGGAACCTGTATGAAGAGCTCCAAGAAT	4006
QY		295	TCTGCAACGAACACAACACAGCACCTCTTTCTGTCGTTCTTCTAGCTGGCTTCCTGCGCGTC	354
Db		4007	TCGTAAAGTAGAGCAAAATGACGCCCTTTCGCCGTGTGTAGGGGCTTCGCGCGCAACC	4066
QY		355	ATTATCGTCTCAGAGCTGTTGAAGAGCGCTCATTTGTTACACCAATTTGCCAATCGCAACC	414
Db		4067	ATTATCGTCTACCGGGCGGAGACTCGATCATCGCAGCCCAATCGCGAACCAGCAACC	4126
QY		415	GACCTGAACCTGGAGGATATCATCGCGTCTTTCTCAATAGCGAGTGTATGCGAATCAACA	474
Db		4127	GCCAGGAGCTTGAAACATGATCGGCTTCTTCGTCAACACCAATCATSCGAATCACGG	4186
QY		475	TAGATCATCAGGATACCTTTTGGGACTTTGATCAACCAAGTCAAGGCTACGACGACAGCAG	534
Db		4187	TCGACGGCGACGACACTTTTGAAGCGTGGTGGCGCAAGTTCGGAGCACCGCGACGCGG	4246
QY		535	CATTGAGACGAGGATATTCCTTTTCAGCGCGTTGTATCAGCACTACAGCCTTGGATCCA	594
Db		4247	CATTGAGACCAAGAGCTGCCCTTTGAGCGCGTGTGAGGGCACTCCTTCCACGCTCGA	4306
QY		595	GAGATCTGTCAAGCACACCTCTCGCAACACTCATTTTGCAGTGCACCTACAGAAGSACC	654
Db		4307	GAGACCTATCCGAAACCACTAGCACAGCTACCTTCGCTCTCATCTCACAGGACC	4366
QY		655	TTGGAAGATTCAGTTCCAGGGTCTCGAGTCGTAACCTGTGCCTAGCAAGCGTACACTC	714
Db		4367	TCGGCAAGTTCGAGCTGGAGGGTCTCGTAGCGAAACCCGCTCGAACAAGGTATACCCA	4426
QY		715	GATTTGACATGGAGTTCATCTGTTTCAAGAAACCGACAGCCTTAAAGTAGCGTCAACT	774
Db		4427	GGTTCGACGTGGAGTTTCACTCTGTCCAAAGACCGGAAGACTTAAGCGGTAAACGTGGCAT	4486
QY		775	TTCCGATGAGCTGTTCAAATGSGAGACTGTTCAAATGTCGTCAGAGTATTTCTTTCAGA	834
Db		4487	TTCCGCGATCATTTATTCAGCCTTGAGACCAATTAGCAATGTAGTCGCCATTTTTTCCAAA	4546
QY		835	TTCTGAGAAACGGGCTTCAAAGTTCGCGGACACAGTCTCAATACTTCTTTTCACATGATG	894
Db		4547	TCCTGGACAAAGCATATGCCACGCTTCGACTCCAATCGTGTCTCTCCGCTTTACCGATG	4606
QY		895	GCATTTGACTCTTGAAAAATTTGGATGTTCTCAACGTCFAAACATGTGCACTATCCCCGAG	954
Db		4607	GGTTAGCGGACCTTCGTGCCATGGGCTTGCTTCAGATCGAAGAGGAGAATAACCCGCGG	4666
QY		955	AATCGAGCTGGCTGATGCTTCCAGACCAAGTCTCTGCTTACC	1000
Db		4667	AGTCGAGCTCTGTCAGCTGTTCCGAAGCAGAGTGGCGCTCAACC	4712

RESULTS 3

RESULT 3
AAO54386

AAQ54386
ID AAQ54386 standard; DNA; 46899 BP.

XX
AC

DT 08-JUL-1994 (first entry)

XX	T. niveum Cyclosporin synthetase gene.
DE	
XX	Enzyme: cyclosporin; synthetase-like activity; Tolypocladium niveum;
KW	T. inflatum GAMS; biosynthesis; vector; cyclosporin synthetase; ss.
KW	
XX	Tolypocladium niveum.
XX	
XX	Key Location/Qualifiers
PH	CDS 885..46730
FT	/*tag= a
FT	/product= Cyclosporin synthetase
FT	40239..43129
FT	/*tag= b
FT	/note= "Sali restriction fragment, preferred
FT	fragment, Claim 4"
FT	
FT	misc_feature 37781..40244
FT	/*tag= d
FT	/note= "Sali restriction fragment, preferred
FT	fragment, Claim 5"
XX	
XX	EP578616-A.
PN	
XX	12-JAN-1994.
PD	
XX	
XX	05-JUL-1993; 93EP-0810474.
XX	
XX	09-JUL-1992; 92AT-0001403.
PR	08-MAR-1993; 93AT-0000437.
PR	29-APR-1993; 93CH-0001310.
PR	04-MAY-1993; 93CH-0001375.
XX	
XX	(SANO) SANDOZ LTD.
PA	(SANO) SANDOZ PATENT GMBH.
PA	(SANO) SANDOZ-ERFINDUNGEN VERW GES MBH.
XX	
XX	Leitner E, Schneider E, Schoergendorfer K, Weber G;
PI	
XX	
DR	WPI: 1994-010432/02.
DR	P-PSDB; AAR44929.
XX	
PT	Isolated DNA sequence - which codes for enzyme having cyclosporin
PT	synthetase like activity
XX	
PS	Claim 6; Page 17-41; 93pp; English.
XX	
CC	This sequence encodes an enzyme which has cyclosporin synthetase-
CC	like activity. This sequence was isolated from Tolypocladium niveum
CC	(formerly known as T. inflatum GAMS). The enzyme encoded by this
CC	sequence catalyses the peptide biosynthesis of cyclosporins and
CC	structurally related molecules. This sequence may be used for the
CC	production of cyclosporin by transforming a vector containing this
CC	sequence in to a recombinant host. This allows effective production
CC	of antibiotic cyclosporin or its derivatives.
XX	
SQ	Sequence 46899 BP; 10651 A; 13513 C; 12509 G; 10226 T; 0 other;

	Query Match	33.9%	Score 339.4	DB 15	Length 46899
	Best Local Similarity	60.1%	Pred. No. 1.1e-98		
	Matches 603	Conservative 0	Mismatches 391	Indels 9	Gaps 2
Qy	4	TTGATGCTTTGCAGCGGATCTCAAATCAGCTCTACTCAGCTGCGCTCAAGSAGCTCAAAG	63		
Db	30226	TTGATATCTTGAGACAGGAGCTCGGTCAAATCTACTCAGCGCTTTTACGTGGC----	AGGG	30288	
Qy	64	ACCCGGTGTGCAGCACTACATCCCTCTACCTATCCAGTACAGCGACTTTTGCAAAATGGCAGA	123		
Db	30283	ATCCGTTATCTCAGGTCAAGCCCTCCCAATACAATATCTGACTTTGCGGCTTGGCAGA	3034		
Qy	124	AGGA-----CCAAATTCATAGACGAGGAGCAACTCACTACTGGAAGSAGCAACTCA	177		
Db	30343	AGGAAGTGCCTCAAGTTGCCGACGTAGAGGAGCTGCGGTACTGCGGAGAACCCAGTTAG	3040		


```

Db 404 GCTATAGCGGCGAGTCGACATTCGCTGGGGTACCTATCGCAACCGCAACCGCGCGC 463
QY 421 AACTGGAGGATATCATCGCTGCTTGTCAATAGCGAGTATGCGAATCAACATAGATC 480
Db 464 AGGTCGAGCGCTGATCGCTTCTTCGTCATATCCAGGTCGCTGCTTGCAGGTCGATG 523
QY 481 ATCAGGATACCTTTGGGACTTTGATCAACCAAGTCAAGGCTACGACGACGACGATTCG 540
Db 524 CTGGCTTGGCTTTCGGGATCTACTTGGGCGCGCTGCGAGGCGCGCTGGCGCGCAGG 583
QY 541 AGAAGAGGATATTCGCTTTCAGCGCGTGTATCAGCAGCTACAGCCTGGATCAGAGATC 600
Db 584 CGCACCAGGATCTCGCTTTCAGCAATGTGTCGATGCTTGCAGCCCGCAAGCGCA----A 638
QY 601 TGTCAAGCACACCTCTCCGCAACCTATTTTTCAGTGCATCTACAGAGGACCTTGGAA 660
Db 639 TCTCAGCCACAGCCCGTGTTCAGGTGATGTATTAACCAAGAGCGCGAGCGCGCAGGA 698
QY 661 GATTCAAGTTCAGGCTCTCGAGTCCGTACCTGTCCTAGCA--AAGCGTACACTCGATT 718
Db 699 TGCCCAAGTCGATGTTGACATCTCGAGAGTTTTCCTGGGATGTTGCTGCCGACAGTT 758
QY 719 TGACATGGAGTTCATCTGTTTCAAGAAACCGAGCGCTTAAAGGTAGCGTCAACTTGC 778
Db 759 CGATCTTGGCTCGATACCTGGGAACCCCGGAGCGCTTGGGCGCGCTGACCTACGC 818
QY 779 CGATGAGCTGTCAAAATGGAGATGTTGAAATGTCGTAGAGTATCTTTTGATGATCT 838
Db 819 GACCGAGCTGTTCAGGCGGCGACCGTCGAGCGCATGGCGCGCATTTGGCAGAACCTGCT 878
QY 839 GAGAAACGGCTTCAAAAGTTCGCGACACAGTCTCAATACTTCTTCTGACTGATGCAT 898
Db 879 CGCGGCGATCTGGAACACCGCGAGCGCGTCTGCTGCGGATGCTGCGATGCCGA 938
QY 899 TGTGACTCTTGAATAATTTGATGTTCTCAACGCTCAAAATGTCGACTATCCCGAGATC 958
Db 939 GGAGCGTGGCAGTGTCTGGAAGCTGGAAGCGCACTGCGCGGAGTACCGCTGCAACG 998
QY 959 GAGCTTGGCTGATGTTCTTCAGACCCCAAGTC 989
Db 999 CGCGTGCACCGGTTGTTTCGAGGAGCAGTC 1029

```

RESULT 9

AAA58472
ID AAA58472 standard; DNA; 18660 BP.

XX AAA58472;

XX 31-OCT-2000 (first entry)

DE Nucleotide sequence of the bleomycin (BLM) gene cluster ORFs 31-40.

XX BLM gene cluster; bleomycin gene cluster; polyketide metabolite;
KW bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;
KW thiazoline; bithiazoline; microbial metabolite; sugar; ss.

OS Streptomyces verticillus.

XX Key Location/Qualifiers

```

FH 1..354
FT CDS
FT /*tag= a
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF31; encodes AAB07580"
FT 925..9450
FT FT
FT /*tag= b
FT /note= "ORF32; encodes AAB07581"
FT 9447..10802
FT FT
FT /*tag= c
FT /note= "ORF33; encodes AAB07582"
FT 10864..11877
FT FT
FT /*tag= d

```

```

FT /note= "ORF34; encodes AAB07583"
FT 11888..12640
FT /*tag= e
FT /note= "ORF35; encodes AAB07584"
FT 12637..13920
FT /*tag= f
FT /note= "ORF36; encodes AAB07585"
FT 13928..14746
FT /*tag= g
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF37; encodes AAB07586"
FT 14743..16479
FT /*tag= h
FT /note= "ORF38; encodes AAB07587"
FT 16476..17462
FT /*tag= i
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF39; encodes AAB07588"
FT 17646..18659
FT /*tag= j
FT /note= "ORF40; encodes AAB07589"
FT /note= "no termination codon given"

```

XX WO2000040704-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US000445.

XX 06-JAN-1999; 99US-0115435.

XX 05-FEB-1999; 99US-0118848.

XX 05-JAN-2000; 2000US-0477962.

XX (REGC) UNIV CALIFORNIA.

XX Shen B, Du L, Sanchez C, Chen M, Edwards DJ;

XX WPI; 2000-465974/40.

XX P-PSDB; AAB07580, AAB07581, AAB07582, AAB07583, AAB07584, AAB07585,

XX AAB07586, AAB07587, AAB07588, AAB07589.

XX New bleomycin gene cluster components useful for peptide and/or
XX polyketide metabolites, especially bleomycin, production and for
XX chemically modifying biological molecules -

XX Claim 8; Page 137-153; 162pp; English.

XX The present sequence represents the BLM (Bleomycin) gene cluster,
XX containing open reading frames (ORFs) 31-40. The proteins encoded
XX by the gene cluster are useful for producing peptides and/or polyketide
XX metabolites, especially bleomycin or bleomycin analogues. They are
XX also useful for chemically modifying biological molecules to produce
XX branched methyl groups, and for coupling amino acids and fatty
XX acids. They may be reacted with an apo-carrier protein and coenzyme A
XX to produce a holo-carrier protein. The BLM gene cluster or catalytic
XX domains can be used individually or collectively to produce
XX thiazolidine, thiazoline, bithiazoline and bithiazoline-containing
XX microbial metabolites. The BLM gene cluster may also be used to produce
XX sugars.

XX SQ Sequence 18660 BP; 2365 A; 7239 C; 6510 G; 2546 T; 0 other;

XX Query Match 7.4%; Score 74.2; DB 21; Length 18660;

XX Best Local Similarity 49.1%; Pred. No. 1.6e-12;

XX Matches 258; Conservative 0; Mismatches 258; Indels 9; Gaps 2;

XX QY 73 CAGCACTCACTCTCTACCTATCCAGTACGACGACTTGCAGTCCAGAACGACCAAT 132

XX Db 5642 CCGGGCTCCCTCCGCTCGAGTCCAGTCCGCGCTCTGCGACGCGTCTCTGGA 5701

XX QY 133 T-----CATAGACGAGGAGCAACTCAACTACTTGGAGAGCAACTCAAGACTCTTT 186

XX Db 5702 TGACCGGCCCGCTGCGGAGGAGCACCTCGCGTACTGGAAGCGGCCCTCGACGGGCAC 5761

(GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA (DRFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
XX
PI Fraser CM, Tuemmler B, Hohelsel J, Duesterhoeft A, Hilbert H;
PI Timmis KN, Moore E, Straetz M, Heim S, Golyshin P;
XX WPI; 2001-159715/16.
XX
XX New DNA encoding a bacterial lipopeptide synthase, useful e.g. for
PT producing antimicrobial agents, surfactants with increased
PT bioavailability in microbial depollution methods, and transgenic plants
XX
XX
PS Claim 3a; Page 28-30; 33pp; German.
XX
XX This invention describes a novel DNA sequence (I) that expresses products
CC having the biological function of lipopeptide synthase. The invention
CC also describes (1) recombinant expression vectors containing (1); (2)
CC prokaryotic and eukaryotic cells transformed or transfected with (1) or
CC the vector of (2); (3) production of lipopeptide synthases by culturing
CC cells of (2); (4) expression products (II) of (1), and synthetic proteins
CC or peptides with the same sequences; (5) mono- or poly-clonal antibodies
CC (Ab) specific for (II); (6) hybridoma cells that produce monoclonal Ab;
CC and (7) transgenic plants that contain cells of (2). (1), and their
CC fragments, are useful for expression of recombinant lipopeptide
CC synthases, and as probes and primers for detection, isolation and
CC amplification of full-length cDNA sequences. (1) are used to produce
CC transgenic plants. Lipopeptide synthases are useful for production of
CC antiviral, antibacterial or antifungal lipopeptides, particularly useful
CC for treatment of plants, and for production of lipopeptide surfactants
CC used for increasing bioavailability in microbial depollution processes
CC (acting as emulsifiers) and in tertiary crude oil recovery. Individual
CC subunits from different lipopeptide synthases may be combined to allow
CC synthesis of many different biologically active substances.
XX
SQ Sequence 5451 BP; 937 A; 1766 C; 1807 G; 941 T; 0 other;

Query Match 7.08; Score 70.4; DB 22; Length 5451;
Best Local Similarity 48.24; Pred. No. 1.3e-11;
Matches 265; Conservative 0; Mismatches 276; Indels 9; Gaps 2;

Qy 48 CTCAGGACTCAAAAGACCGCTGTACAGCACTACTCTCTACCTATTCAGTACAGCGAC 107
Db 874 CACCATCCCGCAGGGGAGCGAGCCGATTTCCGCGCTTTCGGTGCATACCGCGAC 933
Qy 108 TTTCGAAATGGCAGAGACCAATTCATAGACGAGGAGCAAA-----CTCAACTAC 161
Db 934 TATGCCAGCTGGCAACGCGAGCGCTGAGCGAAGGCCAGATGCAAGCCAGCTCGGCTAC 993
Qy 162 TGGAGAGCACTCAAGACTCTTCCCGAG---CAAAGATCCCGACCGACTTGGCCCG 218
Db 994 TGGCAGCGCCACTTGGAGAGATGATTTCCGAGTGTGTCAGTACCTCGCCGACCGCCACG 1053
Qy 219 CCTGCACCTTGTCTGGAGACGCGAGTTGCGTATCATCTTACCATCGACGCGAGCTCTAC 278
Db 1054 CCCGAGCTGCAAGCTATCGGGGGGAGCTGTGGACGTGCGCTGCGCGCGAGTTGACG 1113
Qy 279 CAGTCCCTTCGAGCCTTCTCAACGAAACAAACACGACCTCTTTTCGTCGTTCTTAGCT 338
Db 1114 GCGGACCTGCGCGCTGGCAGTGCAGTGCAGCCACGCTGTTCCAGCTCTCTCTAGCC 1173
Qy 339 CGGTTCCGTCGCTCATATTCGTCACAGCTGTGTAAGACGCTGTATTTGGTACACCA 398
Db 1174 TCGTTGCGCTGTGCTGGCGCGCTACAGCGGCAACGAGAAGATCAACATCGCGTGC 1233
Qy 399 ATTGCGAATCGCAACCGACCTGAACCTGGAGGATATCATCGGCTGTTTGTCAATACGCA 458
Db 1234 ATGACCAACCGCAATCCCTGGAGCTGGAAGGTTGATCGGGTCTTCTGTCACACCGTG 1293
Qy 459 TGRATGGAATCAACATGATATCATACGATACCTTTTGGGACTTTGATCAACCACTCAAG 518
Db 1294 GTGCTGCGGCTGAGTGTGGACTACAGCGTTCAGTGTTCGAGCAGTGTGTCGCCCA 1353

Qy 519 GCTACGACGACGACGACGACGATTCGAGACGAGGATATTCGTTTGGCCGCTTGTATCAGCA 578
Db 1354 GAAATCAGCTGTCAGGCGCCAGGCAACGAGGACCTGCCATTCGACGCTCTGTCGAGCG 1413
Qy 579 CTACGAGCTG 588
Db 1414 CTGACGCCAG 1423
RESULT 12
AA14651
ID AA14651 standard; DNA; 77536 BP.
XX
XX AA14651;
XX
XX 08-AUG-2000 (first entry)
XX Nucleotide sequence of the FK-520 biosynthetic gene cluster.
XX
XX FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW Streptomyces hygroscopicus var. ascomyceticus; immunophilin;
KW FK-506 binding protein; polyketide compound; transplant rejection;
KW graft-versus-host disease; uveitis; alopecia universalis;
KW autoimmune chronic active hepatitis; inflammatory bowel disease;
KW multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW neurite outgrowth; nerve regrowth; Parkinson's disease;
KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW peripheral neuropathy; ss.
XX
XX Streptomyces hygroscopicus.
XX
FH Key Location/Qualifiers
FT CDS complement (412..1836)
FT FT /*tag= a
FT FT /note= "fkbw gene"
FT FT /*tag= b
FT FT complement (2020..3579)
FT FT /note= "fkbv gene"
FT FT 3969..4496
FT FT /*tag= c
FT FT /note= "fkbR2 gene"
FT FT complement (4595..5488)
FT FT /*tag= d
FT FT /note= "fkbR1 gene"
FT FT 5601..6818
FT FT /*tag= e
FT FT /note= "fkbE gene"
FT FT 6808..8052
FT FT /*tag= f
FT FT /note= "fkbF gene"
FT FT 8156..8824
FT FT /*tag= g
FT FT /note= "fkbG gene"
FT FT complement (9122..9883)
FT FT /*tag= h
FT FT /note= "fkbH gene"
FT FT complement (9894..10994)
FT FT /*tag= i
FT FT /note= "fkbI gene"
FT FT complement (10987..11247)
FT FT /*tag= j
FT FT /note= "fkbJ gene"
FT FT complement (11244..12092)
FT FT /*tag= k
FT FT /note= "fkbK gene"
FT FT complement (12113..13150)
FT FT /*tag= l
FT FT /note= "fkbL gene"
FT FT complement (13212..23988)
FT FT /*tag= m
FT FT /note= "fkbC gene"
FT FT complement (13452..13662)
FT misc_feature


```

Db 48334 GCCGGCCCTGCCGAACCTGCCCGCTTGCCGGTGCAGTACGCCGACTTCGCCCGCTGGAG 48393
Qy 123 AAGACCAATTC-----TAGACGAGGAGAGCACTCACTACTGGAAGAACAATC 176
Db 48394 CGGCGCGAATTCACCGCGCGGCTGACAGCGCTTGGCCCTACTGGCGGAGCAATC 48453
Qy 177 ---AAAGACTTCCCGAGCAAAAGATCCGACCGACTTTGCCCGCTGCACTTCTGTCT 233
Db 48454 CGGGCGCGCGCGCGCTGCCCTCCCAACCGACCGTCCCGCGCGCGGTCCCGGAC 48513
Qy 234 GGAGACCGAGTTCGCTACATGTTACCATCGAGCGGAGTCTTACCACTCCCTTCGAGCC 293
Db 48514 GCGACGCGCGCATGCGCGAGTGGCGCGCGCGCGCTGGCGCACCGCGGTCTCTCAG 48573
Qy 294 TTCTGCAACGAACAACACAGACTCTTTCGTCGTTCTTCTAGCTGCGTTCGCGCGCT 353
Db 48574 CTCGCGCGGACTCGGTGTCGTCGTTCTATGACCGCTGTGGCGCGCTTCCAAGCGGT 48633
Qy 354 CATATCTCTCACAGCTGTTGAAGCGCTGTATGTTGATACCAATTCGGAATCGCAAC 413
Db 48634 CTCGCCCGGAGCGGCGCACGCGGAGTCTGCTGGTGGCGACCGCTGGCGGACCGTACG 48693
Qy 414 CGACCTGAATGAGGATATCAGCGCTGCTTTGTCATTAATGAGGAGTATGCAATCAAC 473
Db 48694 CGGCGCGCTGACGAGGCGCTGATCGGCATGTTGTCATCAACACGCTGCGCGTCCGCGG 48753
Qy 474 ATAGATCATCAGCATACCTTTGGGACTTTGATCAACCAAGTCAAGCTACGACGACGCA 533
Db 48754 CTCTCGGCGATCGCTGCTTCGCGGACTCTCGACCGTGGCGGCGCACGACCGGAC 48813
Qy 534 GCATTGAGAACGAGGATATTCGTTTGAGCGCGT 568
Db 48814 GCGTTCGCCACGCGACCTGCGGCTTCGAGAACGT 48848

```

RESULT 13

AAL40781 ID AAL40781 standard; DNA; 88421 BP.

XX AC AAL40781;

XX AC AAL40781;

XX AC AAL40781;

XX AC AAL40781;

XX AC AAL40781;

DE 03-OCT-2002 (first entry)

DE 88421nt genomic DNA of ramoplanin producing Actinoplanes sp.

KW Ramoplanin; ramoplanin biosynthetic pathway polypeptide; antibiotic;
 KW biosynthesis gene cluster; bioengineering; peptide synthetase module;
 KW adenylation domain; hydroxyphenylglycine; Hpg; antibiotic precursor;
 KW chlorinate; lipdipeptide; gene; ds.

XX Actinoplanes sp.

XX Key Location/Qualifiers

FT 2077..3078

FT CDS

FT /*tag= a

FT /product= "Protein of ORF 1"

FT 3118..4032

FT CDS

FT /*tag= b

FT /product= "Protein of ORF 2"

FT 4038..5048

FT CDS

FT /*tag= c

FT /product= "Protein of ORF 3"

FT complement (6665..5814)

FT CDS

FT /*tag= d

FT /product= "Protein of ORF 4"

FT complement (7703..6693)

FT CDS

FT /*tag= e

FT /product= "Protein of ORF 5"

FT complement (9464..8130)

FT CDS

FT /*tag= f

FT /product= "Protein of ORF 6"

FT 9691..10761

FT CDS

FT /*tag= g

```

FT CDS /product= "Protein of ORF 7"
FT complement (12751..10829)
FT /*tag= h
FT /product= "Protein of ORF 8"
FT complement (13617..12802)
FT /*tag= i
FT /product= "Protein of ORF 9"
FT complement (15203..13614)
FT /*tag= j
FT /product= "Protein of ORF 10"
FT 15591..15863
FT /*tag= k
FT /product= "Protein of ORF 11"
FT 15880..19035
FT /*tag= l
FT /product= "Protein of ORF 12"
FT 19032..39713
FT /*tag= m
FT /product= "Protein of ORF 13"
FT 39713..65800
FT /*tag= n
FT /product= "Protein of ORF 14"
FT 65826..66530
FT /*tag= o
FT /product= "Protein of ORF 15"
FT 66546..67370
FT /*tag= p
FT /product= "Protein of ORF 16"
FT 67384..70059
FT /*tag= q
FT /product= "Protein of ORF 17"
FT 70099..70662
FT /*tag= r
FT /product= "Protein of ORF 18"
FT 70659..71906
FT /*tag= s
FT /product= "Protein of ORF 19"
FT complement (73439..71964)
FT /*tag= t
FT /product= "Protein of ORF 20"
FT complement (74216..73563)
FT /*tag= u
FT /product= "Protein of ORF 21"
FT complement (75424..74213)
FT /*tag= v
FT /product= "Protein of ORF 22"
FT 75535..76464
FT /*tag= w
FT /product= "Protein of ORF 23"
FT complement (78110..76449)
FT /*tag= x
FT /product= "Protein of ORF 24"
FT complement (79864..78107)
FT /*tag= y
FT /product= "Protein of ORF 25"
FT complement (81624..79861)
FT /*tag= z
FT /product= "Protein of ORF 26"
FT complement (81909..81682)
FT /*tag= aa
FT /product= "Protein of ORF 27"
FT complement (82346..82062)
FT /*tag= ab
FT /product= "Protein of ORF 28"
FT 82587..84446
FT /*tag= ac
FT /product= "Protein of ORF 29"
FT 84481..85548
FT /*tag= ad
FT /product= "Protein of ORF 30"
FT 85536..86845
FT /*tag= ae
FT /product= "Protein of ORF 31"

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: June 4, 2003, 19:04:47 ; Search time 57 Seconds
(without alignments)
5385.674 Million cell updates/sec

Title: US-09-482-788-1_COPY_4000_5000

Perfect score: 1001

Sequence: 1 caattgatgtctgcagcgc.....cccaagtctgtgtacc 1001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCFUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	339.4	33.9	46899	1	US-08-471-119A-1
2	62.8	6.3	178	1	US-08-471-119A-3
3	60.4	6.0	474	2	US-08-403-852D-14
4	60.4	6.0	474	3	US-08-510-646B-14
5	60.4	6.0	474	4	US-09-231-818-14
6	60.4	6.0	4403765	4	US-09-103-840A-2
7	60.4	6.0	411529	4	US-09-103-840A-1
8	53.6	5.4	2219	3	US-08-510-646B-17
9	47.2	4.7	4403765	4	US-09-103-840A-2
10	47.2	4.7	4411529	4	US-09-103-840A-1
11	41.2	4.1	11601	2	US-08-222-617A-3
12	41.2	4.1	11601	2	US-08-222-617A-24
13	33	3.3	330	2	US-08-743-200-5
14	33	3.3	8532	1	US-08-452-655B-1
15	33	3.3	8532	3	US-08-450-582-1
16	33	3.3	9606	1	US-07-741-940-1
17	33	3.3	9606	1	US-08-289-548A-1
18	33	3.3	9606	1	US-08-452-654-1
19	33	3.3	9606	2	US-08-370-235A-1
20	33	3.3	9606	4	US-08-449-731-1
21	32.8	3.3	6827	1	US-08-222-616-17
22	32.8	3.3	6827	4	US-08-446-648-17
23	32.8	3.3	6827	5	PCT-US95-04228-17
24	32.6	3.3	7218	1	US-08-232-463-14
25	32.6	3.3	26385	4	US-08-961-527-3
26	31.8	3.2	5365	4	US-08-961-527-77
27	31.4	3.1	807	1	US-07-706-691G-97

28	31.4	3.1	807	1	US-08-254-021-97
29	31.4	3.1	807	2	US-08-618-446-97
30	31.4	3.1	807	3	US-08-980-135-97
31	31.4	3.1	807	4	US-09-585-798-97
32	31.4	3.1	1395	4	US-09-586-719-3
33	31.4	3.1	3601	4	US-08-334-179A-1
34	31.4	3.1	3603	4	US-08-462-467B-1
35	31.4	3.1	4776	2	US-08-852-401-1
36	31.2	3.1	4195	1	US-08-340-011-1
37	31.2	3.1	4195	3	US-08-901-710-1
38	31.2	3.1	4416	3	US-08-795-430-1
39	31.2	3.1	4416	4	US-09-355-700-1
40	31.2	3.1	4416	4	US-08-601-132-36
41	31.2	3.1	4425	1	US-08-222-616-31
42	31.2	3.1	4425	4	US-08-446-648-31
43	31.2	3.1	4425	5	PCT-US95-04228-31
44	31.2	3.1	4795	1	US-08-340-011-3
45	31.2	3.1	4795	3	US-08-901-710-3

ALIGNMENTS

RESULT 1

US-08-471-119A-1
; Sequence 1, Application US/08471119A
; Patent No. 5827706
; GENERAL INFORMATION:
; APPLICANT: Leitner, Ernst
; APPLICANT: Schneider, Elisabeth
; APPLICANT: Schoergendorfer, Kurt
; APPLICANT: Weber, Gerhard
; TITLE OF INVENTION: Cyclosporin Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 5827706artis Corporation
; STREET: 59 Route 10
; CITY: East Hanover
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07936
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,119A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kassenoff, Melvyn
; REGISTRATION NUMBER: 26,389
; REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 503 8474
; TELEFAX: 201 503 8807
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46899 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Tolypocladium niveum
; STRAIN: ATCC 34921
; US-08-471-119A-1

Query Match 33.9%; Score 339.4; DB 1; Length 46899;
Best Local Similarity 60.1%; Pred. No. 2.2e-101;

Matches	603; Conservative	0; Mismatches	391; Indels	9; Gaps	2;
Qy	4	TTGATGCTTTGGACGCGATCTCAATCAGCTCTACTCAGCTCGCGTCAAGACTCAAAAG	63		
Db	30226	TTGATATCTTGACAGAGGCTTCGCTCAATCTACTCAGCGCTTACGTGGC---AGG	30282		
Qy	64	ACCGCTGTACAGCACTACTCTCTACCTATCCACTACAGCAGTTTGCAAAATGGCAGA	123		
Db	30283	ATCCGTTATCTCAGGTCAAGCCCTCCCAATAACAATATCGTACTTTGCGCTTTGGCAGA	30342		
Qy	124	AGGA-----CCAATTCATAGACGAGGAGAAGCAACTCAACTACTTGGAGAAGCAACTCA	177		
Db	30343	AGGAAGCTGCCCAAGTTGCCGAGCATGAGAGCAGCTCGGTACTGGGAGAACAGTTAG	30402		
Qy	178	AAGACTTTCGCCAGCAAGATCCCGACCGACTTTGCCCGCCCTGCATCTCTGCTGGAG	237		
Db	30403	CTGACAGTACTCCGGTGAGCTTCTGACCGACTTTCCCGGCCACAGTTCCTGAGTGGGA	30462		
Qy	238	ACGCAGTTTCCGTACACTGTACCACTGAGGCGGAGCTCTACCGATCCCTTCGACGCCTTCT	297		
Db	30463	AGGCTGGTGTATCCCGGTACCATTTGAGGGCCGGTCTACGAGAAGCTTCTGAAGTTCT	30522		
Qy	298	GCAACGAACACACACGACTCTTTCGCTGTTTCTTCTAGCTCGCTTCGTCGCCCTCATTT	357		
Db	30523	CCAAGGAGCGCAGGTAACTCTGTTCTCGGTGCTATTAAACAGCTTCGGGCCACACACT	30582		
Qy	358	ATCGTCTCACAGCTGTTGAGACGCTGTCTATTTGGTACACCAATTTGCAATCGCAACGAC	417		
Db	30583	TTTCGTCTCACTGGTCAGAGGATGCTACGATCGGTACCCCAATTTGCAATTCGAACGGC	30642		
Qy	418	CTGAACTCGAGGATATCATCGGCTGCTTTGTTCATACGACGTGTATCGAATCAACATAG	477		
Db	30643	CAGAACTCGAGCATATCATTTGATTTCTGCTCACACCCCAATGATGGCTTCTCTCTCG	30702		
Qy	478	ATCATCACGATACCTTTGGGACTTTGATCAACCAGTCAAGGCTACGACGACGACGAT	537		
Db	30703	ATACGGCAGCACATTCGAAATCCCTTAGTCCAGCATGTTCCGTCGCTGCTACAGATGCCT	30762		
Qy	538	TCGAGAACGAGATATTCGGTTTGAGCGGTTGTATCAGCACTACAGCCTGGATCCAGAG	597		
Db	30763	ATTCCATACGAGTATTCCTTTGCAACGGATCGCTCTCGGCACTTCTCCTTGGCTCGAGAG	30822		
Qy	598	ATCTGTCAAGCACACTCTCGCACAACTCATTTTGTGAGTGCACCTACAGAAGGACCTTG	657		
Db	30823	ATGCCTCAGGAAGCCCACTAATCCAGCTTATGTTTGCTTGCATCTACAGCCAGATCTCG	30882		
Qy	658	GAGATTTCAAGTTCAGGGTCTCAGTCCGTAACCTGTGCTCAGCAAAAGCGTACACTCGAT	717		
Db	30883	GGAACTATTACTTCGAAGGACTCGAGCATGAGCGCTGCCAAACAGGCTGCCAACACGTT	30942		
Qy	718	TTGACATGGAGTTCCTATCTTTCAAGAAACCGACAGCCTTTAAAGGTAGGCTCAACTTG	777		
Db	30943	TCGACATGGAGTTCACCTGTTCGAAGAGCCTACACAGCTGAGTGGTTCAATACTCTTTG	31002		
Qy	778	CCGATGAGCTGTTCAAAATGGAGACTGTTGAAAATGTGCTCAGAGTATTTCTTTGAGATTC	837		
Db	31003	CCGATGAGCTTTCAGAGCTGAAACAAATCAACAGCGCTGCTGACTGTGTTCCAGGAGATAC	31062		
Qy	838	TGAGAAACGGGCTTCAAAAGTTCCGGGACACAGTCTCAATACTTCTCTTCTGACTGATGCA	897		
Db	31063	TCGACGCGGCTTCGACCACCCCAAGTCTCCATTTCTACTATGCCCTGACTGATGGGT	31122		
Qy	898	TTGTGACTTTGAAAATTTGGATGTTCTTCAACGTTCAACCATGTGCAGTATCTCCCGGAGAT	957		
Db	31123	TGATGATCTCGAGAAATCGGCTTGCTTGGAAATCGAGAGCAGCAACTTCCCTCGGACT	31182		
Qy	958	CGAGCTTGGCTGATGTTCTCCAGACCCCAAGTCTCTGCTTACCC-1000			
Db	31183	ACTCGGTTGTCCAGCTCTTCCAGACAGCAGTGGCTGGCAATCC-31225			

0
E
+
C
D
D

RESULT 2
US-08-471-119A-3

```

; Sequence 3, Application US/08471119A
; Patent No. 5827706
; GENERAL INFORMATION:
; APPLICANT: Leitner, Ernst
; APPLICANT: Schneider, Elisabeth
; APPLICANT: Schoergendorfer, Kurt
; APPLICANT: Weber, Gerhard
; TITLE OF INVENTION: Cyclosporin Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 5827706artis Corporation
; STREET: 59 Route 10
; CITY: East Hanover
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07936
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,119A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kassenoff, Melvyn
; REGISTRATION NUMBER: 26,389
; REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 503 8474
; TELEFAX: 201 503 8807
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 178 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: tolypocladium geodes
; US-08-471-119A-3

Query Match 6.3%; Score 62.8; DB 1; Length 178;
Best Local Similarity 61.7%; Pred. No. 3.5e-11;
Matches 100; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 387 ATTGGTACACCAATTCCGAATCCGACCGACCTGGAAGATATCATCGGCTGCTTT 446
Db 9 ATCGGCTCTCCAATTCGCAACAGAAAATCGAGCAGAGCTTATGGCTGTTTT 68

QY 447 GTCAATACCCAGGTATGCGAATCAACATAGATCATCAGATACCTTCGGACTTTGATC 506
Db 69 GTCAATACTAGTGTATGAGACTGCCAGTTACCGATGAGAATATTCGCCAATTTGATT 128

QY 507 AACCAAGTCAAGGCTACGACGACGACGACGATTCGGAACGAG 548
Db 129 GACTGTGTACGAGAGACGTCACCCGAGGCTTGGACCAACCAAG 170

RESULT 3
US-08-403-852D-14
; Sequence 14, Application US/08403852D
; Patent No. 5891695
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Blanche, Francis
; APPLICANT: Crouzet, Joel
; APPLICANT: Jacques, Nathalie
; APPLICANT: Lacroix, Patricia
; APPLICANT: Thibaut, Denis

```

APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
APPLICANT: De Crey-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,852D
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4400
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 474 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..474
OTHER INFORMATION: /product= "Partie du gene Snpb"
US-08-403-852D-14

Query Match 6.0%; Score 60.4; DB 2; Length 474;
Best Local Similarity 47.2%; Pred. No. 4e-10;
Matches 184; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

QY 199 TCCCGACCGACTTTGCCCGCCCTGCACTTCTGTGTGAGACGAGGTGGTACATGTTA 258
DB 35 TGCCCTGCGACCAACCCCGCGCCGCTGCGCACCCACCGCGCGCCGCTCC 94
QY 259 CCATCGACGCGGAGCTCTACAGTCCCTTCGAGCTTCTGCAAGCAACACGACCT 318
DB 95 ACATCGACGCGCGCTCCAGGAGAGCTGACCGGCTCTCCAAAGGCGCTGCGAGCGC 154
QY 319 CTTTCGTCGTCTTCTAGCTGCGTTCGTCGCGCTCATTATCTCTCACAGCTTTGAAG 378
DB 155 TGTTCATGTCGTCAGGCGCGGTGCGCCGCTGTCTACCCGCGCGCGCGGACCG 214
QY 379 ACCTGTCTATTGGTACACCAATTCCGAATCGCAACCGACCTGAACCTGAGGAGTATCATCG 438
DB 215 ACATCCCGCTCGCGACGCGCGCTGCGCGCGCGCGCGCGCGCTGCGAGACCTGTTGG 274
QY 439 GCTGCTTTGTCATACGCACTGTATGCGGAATCAACATAGATCATCAGTACCTTTGGGA 498

DB 275 GCTTCTCTGTCACACCCCTGCTCTGCGCACCCGACACCTCCGCGGACCCACCTTCGCGG 334
QY 499 CTTTGTATACCAAGTCAAGCTACGACGACGACGATTCGAGACGAGGATATTCGCT 558
DB 335 AACTCGTCGCGACCGGTGCGGCGAGTTGACCTCGCGCTACACGACCGAGGACGCT 394
QY 559 TTGAGCGCGTTGTATCAGCACTACAGCCTG 588
DB 395 TCGAAAGCTCGTGAAGAGTCAACCCG 424

RESULT 4
US-08-510-646B-14
; Sequence 14, Application US/08510646B
; Patent No. 6077699
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Blanche, Francis
; APPLICANT: Crouzet, Joel
; APPLICANT: Jacqoux, Nathalie
; APPLICANT: Lacroix, Patricia
; APPLICANT: Thibaut, Denis
; APPLICANT: Zagorec, Monique
; APPLICANT: Debussche, Laurent
; APPLICANT: De Crey-Lagard, Valerie
; TITLE OF INVENTION: Polypeptides Involved In The
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
; TITLE OF INVENTION: Coding For These Polypeptides And Their Use
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/510,646B
; FILING DATE: 03-AUG-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,852
; FILING DATE: 10-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR 93/00923
; FILING DATE: 25-SEP-1993
; APPLICATION NUMBER: FR 92/11441
; FILING DATE: 25-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03806.0054-01000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 474 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: S.pristinaespiralis

```

;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..474
; OTHER INFORMATION: /product= "Partie du gene Smbd"
US-08-510-646B-14

Query Match
Best Local Similarity 6.0%; Score 60.4; DB 3; Length 474;
Matches 184; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

Qy 199 TCCGACGCGCTTTGCCGCGCTTCTGCTGGAGAGCGAGGTTGCGTACATGTTA 258
Db 35 TGCCCTGGGACACCCCGCGCGCGCTGCGCACCCAGCGCGCGCGCGCTTCC 94

Qy 259 CCATCGAGCGGAGCTTACACAGTCCCTTCGAGCGCTTCTGCAAGCAACACACGACCT 318
Db 95 ACATCGAGCGCGCTTCCACAGAGCTGACCGCGCTCTCCAGGCGCTGCGACGACGCC 154

Qy 319 CTTTCTGCTGTTCTTCTAGCTGCGTTCGCTGCGCTCATATTCGCTCAGAGCTGTTGAAG 378
Db 155 TGTTCATGTTGCTCTCAGCGCGGCTGCGCGCGCTCTCACCGCGCGCGCGCGCGCG 214

Qy 379 ACGTGTGCTATGTTGATGACCAATTCGAATTCGCAACCGACCTGAACCTGGAGGATATCATCG 438
Db 215 ACATCCCGTGGAGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 274

Qy 439 GCTGCTTTGTCAATACGAGTGTATGCGAATCAACATAGATCATCAGGATACCTTTGGGA 498
Db 275 GCTTCTTGTCAACACCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 334

Qy 499 CTTTGTATCAACCAAGTCAAGCTACGAGCGAGCGAGGATTCGAGAGGAGGATATTCGTT 558
Db 335 AACTCGTGGCGAGCGTGGCGAGTTCGACCTTCGCGCGCGCGCGCGCGCGCGCGCGCG 394

Qy 559 TTGAGCGCGTGTATGATGACCACTACAGCGCTG 588
Db 395 TCGAAAGCTGCTGAGAGAGTCAACCCCG 424

```

```

RESULT 5
US-09-231-818-14
; Sequence 14, Application US/09231818
; Patent No. 6171846
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Blanc, Francis
; APPLICANT: Crouzet, Joel
; APPLICANT: Jacques, Nathalie
; APPLICANT: Lacroix, Patricia
; APPLICANT: Thibaut, Denis
; APPLICANT: Zagorec, Monique
; APPLICANT: Debussche, Laurent
; APPLICANT: De Crecy-Lagard, Valerie
; TITLE OF INVENTION: Polypeptides Involved In The
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
; TITLE OF INVENTION: Coding For These Polypeptides And Their Use
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/231.818
; FILING DATE:
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/403.852
; FILING DATE: 10-MAY-1995
; APPLICATION NUMBER: PCT/FR 93/00923
; FILING DATE: 25-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/11441
; FILING DATE: 25-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03806.0054-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 474 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: S.pristinaespiralis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..474
; OTHER INFORMATION: /product= "Partie du gene Smbd"
US-09-231-818-14

Query Match
Best Local Similarity 6.0%; Score 60.4; DB 4; Length 474;
Matches 184; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

Qy 199 TCCGACGCGCTTTGCCGCGCTTCTGCTGGAGAGCGAGGTTGCGTACATGTTA 258
Db 35 TGCCCTGGGACACCCCGCGCGCGCTGCGCACCCAGCGCGCGCGCGCTTCC 94

Qy 259 CCATCGAGCGGAGCTTCTTACAGTCCCTTCGAGCGCTTCTGCAAGCAACACACGACCT 318
Db 95 ACATCGAGCGCGCTTCCACAGAGCTGACCGCGCTCTCCAGGCGCTGCGACGACGCC 154

Qy 319 CTTTCTGCTGTTCTTCTAGCTGCGTTCGCTGCGCTCATATTCGCTCAGAGCTGTTGAAG 378
Db 155 TGTTCATGTTGCTCTCAGCGCGGCTGCGCGCGCTCTCACCGCGCGCGCGCGCGCG 214

Qy 379 ACGTGTGCTATGTTGATGACCAATTCGAATTCGCAACCGACCTGAACCTGGAGGATATCATCG 438
Db 215 ACATCCCGTGGAGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 274

Qy 439 GCTGCTTTGTCAATACGAGTGTATGCGAATCAACATAGATCATCAGGATACCTTTGGGA 498
Db 275 GCTTCTTGTCAACACCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 334

Qy 499 CTTTGTATCAACCAAGTCAAGCTACGAGCGAGCGAGGATTCGAGAGGAGGATATTCGTT 558
Db 335 AACTCGTGGCGAGCGTGGCGAGTTCGACCTTCGCGCGCGCGCGCGCGCGCGCGCGCG 394

Qy 559 TTGAGCGCGTGTATGATGACCACTACAGCGCTG 588
Db 395 TCGAAAGCTGCTGAGAGAGTCAACCCCG 424

```

```

RESULT 6
US-09-103-840A-2/C
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.

```


QY 352 TACACCAATTCGGAATCGAACCGACCTGAATCGAGGATATCATCGCTGCTTTGTCAA 451
 Db 113886 ATTTCCCATCGCCGCGCAGCGATCTCTGCGCTGGATAACCTTGTGGGCTTTTGTCAA 113945
 QY 452 TACGAGTGTATGGGAATCAACATAGATCATCAGATACCTTTGGGACTTTTGATCAACCA 511
 Db 113946 CACCTTGTGTGGGGTCAACCTGGCGGTGATCCCGAGCTTCGCCGGAACCTGCTGGGGCA 114005
 QY 512 AGTAAAGCTACGACGACGAGCATTCGAGAACGAGGATATTCGGTTTGTAGCGCTGTGT 571
 Db 114006 GGTGGAGCGCGCAGCTGCGCGCTACGAAATCAAGACGTACCTTTGAGAGTGCTCGT 114065

RESULT 11
 US-08-222-617A-3
 ; Sequence 3, Application US/08222617A
 ; Patent No. 5862879
 ; GENERAL INFORMATION:
 ; APPLICANT: Veenstra, Annemarie E.
 ; APPLICANT: Martin, Juan F.
 ; APPLICANT: Garcia, Bruno D.
 ; APPLICANT: Gutierrez, Santiago
 ; APPLICANT: Barredo, Jose L.
 ; APPLICANT: Von Doehren, Hans
 ; APPLICANT: Palissa, Harriet
 ; APPLICANT: Van Liempt, Henk
 ; APPLICANT: Montenegro, Eduardo P.
 ; TITLE OF INVENTION: A Method for Influencing Beta-Lactam
 ; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
 ; TITLE OF INVENTION: Quantities of ACV Synthetase
 ; NUMBER OF SEQUENCES: 27
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
 ; STREET: 300 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: USA
 ; ZIP: 60606
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/222.617A
 ; FILING DATE: 04-APR-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; REFERENCE/DOCKET NUMBER: 97,157
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 11601 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; ORIGINAL SOURCE:
 ; ORGANISM: Acremonium chrysogenum
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 388..11526
 ; OTHER INFORMATION: /function= "Enzyme"
 ; OTHER INFORMATION: /product= "ACV Synthetase"
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 8050..8052
 ; OTHER INFORMATION:
 ; OTHER INFORMATION: /note= "NNN-GCC, AGU, AGC, UCU, UCC, UCA, or UCG";
 ; OTHER INFORMATION: Xaa=Ala or Ser "

US-08-222-617A-3
 Query Match 4.1%; Score 41.2; DB 2; Length 11601;
 Best Local Similarity 60.9%; Pred. No. 0.0064;

Matches 67; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
 QY 387 ATTGTACACCAATTCGGAATCGAACCGACCTGAATCGAGGATATCATCGCTGCTTT 446
 Db 7129 ATTGTATTCGGTTGGCCACCGTAACCATCGGACTTTGAGTCGGTTGTCGGCTCTTT 7188
 QY 447 GTCAATAGCAGTGTATCGGAATCAACATAGATCATCAGATACCTTTGG 496
 Db 7189 GTCAACTTGTCTCTCGGGTCAACGTGTCTCAGTCGGACATTCATGG 7238

RESULT 12
 US-08-222-617A-24
 ; Sequence 24, Application US/08222617A
 ; Patent No. 5862879
 ; GENERAL INFORMATION:
 ; APPLICANT: Veenstra, Annemarie E.
 ; APPLICANT: Martin, Juan F.
 ; APPLICANT: Garcia, Bruno D.
 ; APPLICANT: Gutierrez, Santiago
 ; APPLICANT: Barredo, Jose L.
 ; APPLICANT: Von Doehren, Hans
 ; APPLICANT: Palissa, Harriet
 ; APPLICANT: Van Liempt, Henk
 ; APPLICANT: Montenegro, Eduardo P.
 ; TITLE OF INVENTION: A Method for Influencing Beta-Lactam
 ; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
 ; TITLE OF INVENTION: Quantities of ACV Synthetase
 ; NUMBER OF SEQUENCES: 27
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
 ; STREET: 300 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: USA
 ; ZIP: 60606
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/222.617A
 ; FILING DATE: 04-APR-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; REFERENCE/DOCKET NUMBER: 97,157
 ; INFORMATION FOR SEQ ID NO: 24:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 11601 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; ORIGINAL SOURCE:
 ; ORGANISM: Acremonium chrysogenum
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 388..11526
 ; OTHER INFORMATION: /function= "Enzyme"
 ; OTHER INFORMATION: /product= "ACV Synthetase"
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 8050..8052
 ; OTHER INFORMATION:
 ; OTHER INFORMATION: /note= "NNN-AGU, AGC, UCU, UCC, UCA, or UCG";
 ; OTHER INFORMATION: Xaa=Ala or Ser "

US-08-222-617A-24
 Query Match 4.1%; Score 41.2; DB 2; Length 11601;
 Best Local Similarity 60.9%; Pred. No. 0.0064;
 Matches 67; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
 QY 387 ATTGTACACCAATTCGGAATCGAACCGACCTGAATCGAGGATATCATCGCTGCTTT 446

APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/450,582
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/452,655
FILING DATE: 25-MAY-1995
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.49964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: DP2.5(APC)
US-08-450-582-1

Query Match 3.3%; Score 33; DB 3; Length 8532;
Best Local Similarity 55.8%; Pred. No. 2.8;
Matches 63; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 880 TTCCTTTGACTGATGCATTTGACTCTTGAAAAATTGGATGTTCTCAACGTCACACATG 939
DB 8112 TTGTTTGGCTGATTATCTTTGAATCTTTAATGTTGGATTTGCCCTTTTCGAACACT 8053
QY 940 TCGACTATCCCGAGAATCGAGCTGGCTGATGCTTCCAGACCCCAAGTCTCT 992
DB 8052 GTCAATCACGGGGAGGATATTACCTGTGGGAGATCTTCAGATCTAGGATTGT 8000

Search completed: June 4, 2003, 21:50:58
Job time : 93 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2003, 20:51:29 ; Search time 170 Seconds
(without alignments)
8240.805 Million cell updates/sec

Title: us-09-482-788-1_copy_4000_5000

Perfect score: 1001
Sequence: 1 caattgatgtcttgacgc.....cccaagtctctgtacc 1001

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 870385 seqs, 699768693 residues

Total number of hits satisfying chosen parameters: 1740770

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA.*

- 1: /cgn2.6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2.6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2.6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2.6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2.6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2.6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2.6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2.6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2.6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2.6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2.6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2.6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2.6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2.6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	75.6	7.6	13029	10	US-09-815-242-4052 Sequence 4052, Ap
2	71.8	7.2	3471	10	US-09-974-300-2167 Sequence 2167, Ap
3	65.8	6.6	88421	9	US-09-976-059-1 Sequence 1, Appli
4	60.4	6.0	7347	10	US-09-815-242-7773 Sequence 7773, Ap
5	56	5.6	7158	10	US-09-974-300-2171 Sequence 2171, Ap
6	47.2	4.7	6455	10	US-09-974-300-2166 Sequence 2166, Ap
7	47.2	4.7	8268	10	US-09-974-300-2169 Sequence 2169, Ap
8	41.6	4.2	693	9	US-10-123-155-406 Sequence 406, App
9	34.8	3.5	558	9	US-09-918-995-31196 Sequence 31196, A
10	34.8	3.5	558	9	US-09-918-995-28625 Sequence 28625, A
11	34.2	3.4	269	10	US-09-923-876-5113 Sequence 5113, Ap
12	33.8	3.4	175561	9	US-10-017-721-3 Sequence 3, Appli
13	33	3.3	10383	10	US-09-960-253-181 Sequence 181, App
14	32.8	3.3	6827	10	US-09-982-610-17 Sequence 17, Appl
15	32.6	3.3	1290	10	US-09-815-242-7834 Sequence 7834, Ap
16	32.2	3.2	2277	9	US-10-189-971-21 Sequence 21, Appl
17	32.2	3.2	3173	9	US-10-189-971-5 Sequence 5, Appli
18	32.2	3.2	3753	9	US-10-189-971-15 Sequence 15, Appl
19	32.2	3.2	4026	9	US-10-189-971-23 Sequence 23, Appl

- Sequence 9, Appli
- Sequence 11, Appli
- Sequence 1, Appli
- Sequence 25, Appli
- Sequence 5115, Ap
- Sequence 9092, Ap
- Sequence 2158, Ap
- Sequence 2158, Ap
- Sequence 234, App
- Sequence 75, Appli
- Sequence 1027, Ap
- Sequence 1, Appli
- Sequence 138, App
- Sequence 1, Appli
- Sequence 1, Appli
- Sequence 31, Appli
- Sequence 45, Appli
- Sequence 4, Appli
- Sequence 312, App
- Sequence 75, Appli
- Sequence 498, App
- Sequence 1682, Ap
- Sequence 1379, Ap
- Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-815-242-4052
; Sequence 4052, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4052
; LENGTH: 13029
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa

US-09-815-242-4052

Query Match 7.6%; Score 75.6; DB 10; Length 13029;
Best Local Similarity 45.6%; Pred. No. 4.5e-14;


```
Query Match          6.6%; Score 65.8; DB 9; Length 88421;
Best Local Similarity 48.5%; Pred. NO. 3.4e-10;
Matches 181; Conservative 0; Mismatches 192; Indels 0; Gaps 0;
```

RESULTS. T. A

```

; LIFE: DNA
; ORGANISM: Pseudomonas aeruginosa
;
; FEATURE:
;
; NAME/KEY: CDS
; LOCATION: (1)...(7347)
US-09-815-242-7773

```

Query Match 6.0%; Score 60.4; DB 10; Length 7347;

Best Local Similarity 49.1%; Pred. No. 4.6e-09;
Matches 160; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 270 GAGCTTACAGTCCCTTCGAGCCTTCTGCAACGAAACACACAGCCTCTTTCGTCGTT 329
Db 811 GATCTGTCCAGGGCTCAGCGGCTGGCCAGCAGAGGGGTACTGCTTCGCCCTG 870

QY 330 CTCTAGCTCGTTCCGTCGCGCTCATATTGCTCTACAGCTGTGTAAGACGCTGTCAAT 389
Db 871 TTGCTGGCTTCTGCTCCAGGCTGCTCTGCTGCTACAGCGGCGGATATCCGTGTC 930

QY 390 GGTACACCAATTGGAATCGCAACCGACCTGAACCTGAGGATATCATCGGCTGTGTC 449
Db 931 GCGTGGCTGCTGCAATCGCAACCGGCTGGAGACCGGCTGATCGGCTTCTTCGTC 990

QY 450 AATACGCGATGTATGCGAATCAACATAGATCATCAGCATACCTTTTGGGACTTTCATCAAC 509
Db 991 AACACCGAGTGTCAAGCGGCGCTGAGCGGTGGATGCTTCGAGAGCTGCTGCC 1050

QY 510 CAAGTCAAGGCTAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 569
Db 1051 CAGCGCGCCCAACCGCGCTGGAGCGCCAGCGCCAGCGCCAGCGCTGCGGTCGAGCAACTG 1110

QY 570 GTATCAGCACTACAGCCTGGATCCAG 595
Db 1111 GTGAGGCTTGCAGCGCGGAGCGCAG 1136

RESULT 5
US-09-974-300-2171
; Sequence 2171, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2171
; LENGTH: 7158
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2171

Query Match 5.6%; Score 56; DB 10; Length 7158;
Best Local Similarity 47.7%; Pred. No. 1.4e-07;
Matches 164; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 243 GGTTCGCTACATGTTACCATCGAGCGGAGCTCTACAGTCCCTTCGAGCCTTCTGCAAC 302
Db 3820 GGTACGATTGATCTCAGCATTTGACGAGCTGCAACAGCTGCAAGCGGCTGTTTCGATTTGTCGCCG 3879

QY 303 GAACACAAACAGCACTTCTTCGCTCTTCTAGCTGCGTTCGCTGCGGCTCATTTATCGT 362
Db 3880 AGACCGGATGACGATGTTTATGATTTTTCAGTGGGCTCTTCCCGCCTACTGACACGG 3939

QY 363 CTCACAGCTTTGAAGACGCTGTCTATGTTACACCAATTCGGAATCGCAACCGACCTGAA 422
Db 3940 CTCGAGCAGGCCATGATATTCGCTCGGAGCCCGATAGCGGGAAGAAACGATGATGCT 3999

QY 423 CTGAGAGATATCATCGGCTGCTTTGTCATACGAGTGTATGCAATCAACATAGATCAT 482
Db 4000 TTAGGGGAAATGTCGAGATTGTTTGTATACGCTTGTGCTTCCAAACAGATACATCGGT 4059

QY 483 CACGATACCTTTGGGACTTTTGATCAACCAAGTCAAGGCTACGAGCAGCAGCATTCGAG 542

Db 4060 AATCCGAGTTTTCGGAGCTTCTCAACAGAGTCAAGAAAGTAAACCTTCGAGCGTATGAG 4119

QY 543 AACGAGATATCCGTTTGGCGGCTTGTATCAGCACTACAGCC 586
Db 4120 CATCAGGACCTTCGTTCCGAGCGCTTGTGAAGTGTCTTAATCC 4163

RESULT 6
US-09-974-300-2166
; Sequence 2166, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2166
; LENGTH: 6465
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2166

Query Match 4.7%; Score 47.2; DB 10; Length 6465;
Best Local Similarity 45.6%; Pred. No. 0.00013;
Matches 166; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

QY 199 TCCGAGCAGCTTTCGCCGCTTCTGCTGAGAGCAGGTTGCGTACATGTTA 258
Db 1534 TGCCGTACGATTTTCGCCGCTTCAAAAGCGAGCTTTGAAGGCGCGCTGCTCTTTG 1593

QY 259 CCATCGACGCGAGCTCTACAGTCCCTTCGAGCCTTCTGCAAGCAACACACGACCT 318
Db 1594 GAATCAATCAAAACTACGCGCAGAGCTTCAAAAGCTGCTTCTGAAGCGATCGACAC 1653

QY 319 CTTTCGCTGTTTCTTAGCTCGCTTCGCTCGCTCATTTATCGTCTCACAGCTGTTGAAG 378
Db 1654 TTTATATGTTCTTGTGCGCGCTTAAATATTTTGTGCTCAAGTATGCTCTCAGGAAG 1713

QY 379 ACCTGCTCATTTGTTACACCAATTGCGAATCGCAACCGACCTGAACTGGAGGATATCATCG 438
Db 1714 ATTTAATCGTCGATGCTGTCGCCGCGGAGAACCCATCGGATCTGCACAACATTCGG 1773

QY 439 GCTGCTTTGTCATACGAGCTGTATGGAATCAACATAGATCATCAGCATACCTTTTGGGA 498
Db 1774 GCATGTTGTCATACGCTGCGCTCGGATCGTCCCGAGGGAAGAAACGTTAAGG 1833

QY 499 CTTTGTATCAACCAAGTCAAGGCTACGACGACGAGCATTCGAGAACGAGGATATTCCTG 558
Db 1834 AATTTTGCAGGAGTCAAGAGACAGTCTTCAGGCTTTTGCAGCAGGACTATCCG 1893

QY 559 TTGA 562
Db 1894 TTGA 1897

RESULT 7
US-09-974-300-2169
; Sequence 2169, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene

140	GCAGAGGAGCAACTCACTACTTGGAGAGCACTCTTCCGCCCAAGAT	199
Qy		
287	GCAGAAAAACAAAAAATACAAACAAAGACGCTCTGCTCTCCCGGGAAGAC	346
Db		
200	CCGACCGCACTTTGCCGCCCTGTCACTCTGTCTGGAGAGCC	241
Qy		
347	CCTGACCGGTACATAGCGCTTGGTGTCTCTGCCAGCCAGCC	388
Db		

```

RESULT 10
US-09-918-995-28625
; Sequence 28625, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28625
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(558)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-28625

```

```

RESULT 11
US-09-923-876-5113
; Sequence 5113, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Ito)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 5113
; LENGTH: 269
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc:feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700456379H1
; NAME/KEY: unsure

```

```

; LOCATION: 149, 190
; OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-5113

Query Match          3.4%; Score 34.2; DB 10; Length 269;
Best Local Similarity 54.5%; Pred. No. 0.45;
Matches 66; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 629 TTTTGCAGTGCACCTACAGAGGACCTTTGGAAAGATTCAAGTTCAGGGTCTCGAGTCCGT 688
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 102 TTTTCACGGTGGAGTTCCCGGACTCGCTGGCACGAGCAGTGTCAAGGNTCTCGAGTCGGT 161

Qy 689 ACCTGTGCCTAGCAAGGCTACACTCGATTTGCATGGAGTTCGATCTGTTTCAAGAAC 748
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 162 ACTTCCGCCAAAGCCTTCGTCCCAAGCTGCCACATGGAGATCGATGATCCGAGGGGC 221

Qy 749 C 749
      |
Db 222 C 222

```

```

RESULT 12
US-10-017-721-3
; Sequence 3, Application US/10017721
; Publication No. US20030096248A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Jeanette
; APPLICANT: Daley, George
; APPLICANT: Bolk, Stacey
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF VASCULAR DISEASE
; FILE REFERENCE: MMI-003
; CURRENT APPLICATION NUMBER: US/10/017, 721
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/317,033
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: US 60/330,248
; PRIOR FILING DATE: 2001-10-17
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 175561
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-017-721-3

```

```

RESULT 13
US-09-960-253-181/c.
; Sequence 181, Application US/09960253
; Patent No. US20020123619A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.556
; CURRENT APPLICATION NUMBER: US/09/960,253
; CURRENT FILING DATE: 2001-09-20
; NUMBER OF SEQ ID NOS: 187
; SOFTWARE: Fastseq for Windows Version 4.0

```

```

; SEQ ID NO 181
; LENGTH: 10383
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9089, 9347, 9453, 9519, 10205
; OTHER INFORMATION: n = A,T,C or G
US-09-960-253-181

Query Match          3.3%; Score 33; DB 10; Length 10383;
Best Local Similarity 55.8%; Pred. No. 12;
Matches 63; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 880 TTCCCTTGACTGATGCATGTGACTCTTTGAAAAATTGGATGTTCTCAACGTCACAACTG 939
Db 8150 TTGTTTGCTGATATCTTTTGAATCTTTATGTTGGATTTGCTTTTCTGAAACACT 8091

QY 940 TCGCATATCCCGAGAAATCGAGCTTGGCTGATGCTTCCAGACCCCAAGTCTCT 992
Db 8090 GTCAATCACCGGGGAGTATTACCTGTGGGAGATCTTCCAGATCTAGGATTGT 8038

RESULT 14
US-09-982-610-17
; Sequence 17, Application US/09982610
; Patent No. US20020146420A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; BENNETT, BRIAN D.
; GOEDDEL, DAVID
; LEE, JAMES M.
; MATTHEWS, WILLIAM
; TSAI, SIAO PING
; WOOD, WILLIAM I.
; TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST ANTIBODIES
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/982,610
; FILING DATE: 17-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/446,648
; FILING DATE: 1996-MAY-23
; APPLICATION NUMBER: 08/222616
; FILING DATE: 04-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 40,378
; REFERENCE/DOCKET NUMBER: P0821P3PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6827 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:

```

```

US-09-982-610-17

Query Match          3.3%; Score 32.8; DB 10; Length 6827;
Best Local Similarity 52.1%; Pred. No. 11;
Matches 73; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 348 GCCGCTCATTTATGCTCTCACAGCTGTTGAAGACGCTGTTCATTTGTTCAATACGAGTGTATGCGA 407
Db 2486 GTCACTCTCTGGTGTGTACACCTTGTCTGAAGATGCTTTCAGGGGCCATCCACTTTCAGGGG 2545

QY 408 CGCAACCGGACCTGAAGTGGAGGATATCATCGGCTGCTTGTTCATACGAGTGTATGCGA 467
Db 2546 CAGCCGGGACCTGCCCTTGGGACGTAGTCGGGGTCTTGTGTAGATGTCCCGGCAAGGCC 2605

QY 468 ATCAACATAGATCATCACGA 487
Db 2606 AAGTCACAGATCTTCACCA 2625

RESULT 15
US-09-815-242-7834/c
; Sequence 7834, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7834
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1290)
US-09-815-242-7834

Query Match          3.3%; Score 32.6; DB 10; Length 1290;
Best Local Similarity 48.2%; Pred. No. 4.4;
Matches 92; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 199 TCCCGACCGACTTTTGGCCCGCTGCACCTTCTGTCTGAGACGCGAGTGTGCTACATGTTA 258
Db 782 TCCAGGTCTGACTTGGCGTCTCTTGAAGAACTCGCTGGAGCGAGTCCAGGGCCAGGGTC 723

QY 259 CCATCGACGCGGAGCTCTACCACTCCCTTCGAGCCCTTCTCAACGACACACACGACCT 318
Db 722 ACGTCTCGGCCCCAGCTTGTAGCCGGGCTTCTTCACGCGCTTCGGCGATGCGCGCAGGGCG 663

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2003, 17:53:27 ; Search time 1708.5 Seconds
(without alignments)
9488.839 Million cell updates/sec

Title: US-09-482-788-1_COPY_4000_5000
Perfect score: 100
Sequence: 1 caattgatgttgcagcgc.....cccaagtctgttacc 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum Match 100%
Listing first 45 summaries
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	66.4	6.6	438	17 AQ989462	AQ989462 Rfc000006
2	41.8	4.2	637	17 AQ990256	AQ990256 Rfc000996
3	40.4	4.0	797	17 AQ991640	AQ991640 Rfc01385F
4	38.6	3.9	1145	12 BF036321	BF036321 601459544
5	38	3.8	786	10 BF130762	BF130762 L48-973T3
6	37.8	3.8	791	13 BM012777	BM012777 603637492

c	7	37.2	3.7	492	12	BF653686	BF653686 277599 MA
c	8	37.2	3.7	493	12	BF653730	BF653730 277659 MA
c	9	36.8	3.7	280	12	BF643966	BF643966 NF010A02E
	10	36.8	3.7	483	12	BF646516	BF646516 NF076F01E
	11	36.8	3.7	586	12	EG453877	EG453877 NF095E06L
	12	36.8	3.7	662	12	BF644804	BF644804 NF042H10E
	13	36.8	3.7	668	10	AW689867	AW689867 NF025C03S
c	14	36.6	3.7	591	9	AI737351	AI737351 606039D04
	15	36.6	3.7	638	10	AW056085	AW056085 660004A04
	16	36.4	3.6	939	17	CNS00CNG	AL059400 Drosophil1
c	17	36.2	3.6	816	10	BE641516	BE641516 Cri12.3.H2
c	18	36.2	3.6	840	10	BE640820	BE640820 Cri12.1.H0
	19	35.6	3.6	383	17	BH475809	BH475809 BOHDE17TR
	20	35.6	3.6	736	17	BH179075	BH179075 O13_D_05-
	21	35.6	3.6	761	17	CNS07L8T	AL616031 T3 end of
	22	35.6	3.6	761	17	BH557436	BH557436 BOGNK42TF
	23	35.4	3.5	208	14	BM791536	BM791536 K-EST0071
c	24	35.4	3.5	384	14	T65025	T65025 YC75C05.S1
c	25	35.4	3.5	600	14	BQ783688	BQ783688 fab32f08.
c	26	35.4	3.5	1101	17	CNS00LT2	AL078714 Drosophil1
c	27	35.2	3.5	729	17	BH679719	BH679719 BOHVX54TF
c	28	35.2	3.5	748	17	AG090994	AG090994 Pan trogl
c	29	35.2	3.5	991	12	BG821150	BG821150 602724546
	30	35	3.5	496	17	A2323994	AZ323994 1M0045A18
	31	35	3.5	565	10	AW562597	AW562597 660066H04
c	32	35	3.5	844	10	BE640899	BE640899 Cri12.1.L0
c	33	35	3.5	938	14	BQ619362	BQ619362 RNOSEQH0
	34	35	3.5	1031	11	AY104862	AY104862 Zea mays
	35	34.8	3.5	248	12	BG050913	BG050913 FM1.72.G0
c	36	34.8	3.5	295	9	AA627760	AA627760 nq48C02.S
	37	34.8	3.5	340	9	AI586493	AI586493 486051H08
	38	34.8	3.5	638	10	AW958464	AW958464 EST370534
	39	34.8	3.5	646	10	AW953398	AW953398 EST365468
	40	34.8	3.5	993	14	BQ064890	BQ064890 AGENCOURT
c	41	34.8	3.5	1101	17	CNS00PXE	AL071370 Drosophil1
c	42	34.6	3.5	674	9	AI491368	AI491368 486037H10
c	43	34.6	3.5	675	9	AL692518	AL692518 AL692518
c	44	34.6	3.5	693	9	AL450949	AL450949 AL450949
c	45	34.6	3.5	709	12	BG593086	BG593086 EST491764

ALIGNMENTS

RESULT 1
AQ989462 438 bp DNA linear GSS 14-AUG-2000
LOCUS Rfc000006 Photorhabdus luminescens strain W14 M13 library
DEFINITION Photorhabdus luminescens genomic clone PG000006, DNA sequence.
ACCESSION AQ989462
VERSION AQ989462.1 GI:9648056
KEYWORDS GSS.
SOURCE Photorhabdus luminescens.
ORGANISM Photorhabdus luminescens.
REFERENCE 1 (bases 1 to 438)
AUTHORS ffrench-Constant, R.H., Waterfield, N., Burland, V., Perna, N.T., Daborn, P.J., Bowen, D. and Blattner, F.R.
TITLE A genomic sample sequence of the entomopathogenic bacterium Photorhabdus luminescens W14: potential implications for virulence
JOURNAL Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
MEDLINE 20378633
COMMENT Contact: ffrench-Constant RH
Department of Biology and Biochemistry
University of Bath
South Building, Bath BA2 7AY, UK
Tel: (44) 1225 826621
Fax: (44) 1225 826779
Email: bssrfc@bath.ac.uk
This is one of 2,122 random reads from the M13 library. For annotation of identified clones (BLASTX, BLASTN and mapping to E. coli K12 genome) please see ffrench-Constant et al. 2000, Nucleic

Acids Res.

Seq primer: M13 Forward

Class: shotgun.

Location/Qualifiers

FEATURES

source

1..438

/organism="Photorhabdus luminescens"

/strain="W14"

/db_xref="taxon:29488"

/clone_lib="PLG00006"

/clone_lib="Photorhabdus luminescens strain W14 M13 library"

/dev_stage="primary phase variant"

/note="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janus."

BASE COUNT 92 a 93 c 149 g 104 t

ORIGIN

Query Match 6.6%; Score 66.4; DB 17; Length 438;

Best Local Similarity 55.1%; Pred. No. 6.8e-09;

Matches 130; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 351 GCTCATATCGTCTCACAGCTGTTGAAGACGCTGTCATTGGTACACCAATTGCGAATCGC 410

Db 7 GTTCTGCGCTGTCGGTCAAGAGATATGTCGCGCACACCGAGTCGGGTCCG 66

QY 411 AACCGACTGAATCGAGGATATCATCGGCTGTTGTCAATACGCGAGTGTATGCGAATC 470

Db 67 AATCGTCAGGAAGTAGAACCAATGATGGCTTCTTTGTCATACGCTGGCGTCATATA 126

QY 471 AACATAGATCATACGATACCTTTGGGACTTTGATCAACCAAGTCAAGGTCACGACGACA 530

Db 127 GATGATCGGATGAGCTGATGTAACGAGTTGTCGGCGCAATCCGGCAAAACCGCATG 186

QY 531 GCAGCATTCGAGAACGAGGATATCCGTTTGAGCGCGTGTATCAGCACTACAGCC 586

Db 187 GCGCGCAGGATCAGATACGATTACCGTTTGACAGGTGTGAGATCGTACACACC 242

RESULT 2

LOCUS

AO990256

DEFINITION

Photorhabdus luminescens strain W14 M13 library

Photorhabdus luminescens genomic clone PLG00996, DNA sequence.

ACCESSION

AO990256

VERSION

AO990256.1 GI:9648850

KEYWORDS

GSS.

SOURCE

Photorhabdus luminescens.

ORGANISM

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

Photorhabdus.

REFERENCE

1 (bases 1 to 637)

AUTHORS

ffrench-Constant, R.H., Waterfield, N., Burland, V., Perna, N.T.,

Daborn, P.J., Bowen, D. and Blattner, F.R.

TITLE

A genomic sample sequence of the entomopathogenic bacterium

Photorhabdus luminescens W14: potential implications for virulence

Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)

JOURNAL

MEDLINE

COMMENT

Contact: ffrench-Constant RH

Department of Biology and Biochemistry

University of Bath

South Building, Bath BA2 7AY, UK

Tel: (44) 1225 826621

Fax: (44) 1225 826779

Email: bssrfce@bath.ac.uk

This is one of 2,122 random reads from the M13 library. For

annotation of identified clones (BLASTX, BLASTN and mapping to E.

coli K12 genome) please see ffrench-Constant et al. 2000, Nucleic

Acids Res.

Seq primer: M13 Forward

Class: shotgun.

Location/Qualifiers

1..637

/organism="Photorhabdus luminescens"

source

1..797

/organism="Photorhabdus luminescens"

/strain="W14"

/db_xref="taxon:29488"

/clone_lib="PLG01385F"

/clone_lib="Photorhabdus luminescens strain W14 M13 library"

/dev_stage="primary phase variant"

/strain="W14"

/db_xref="taxon:29488"

/clone_lib="PLG00996"

/clone_lib="Photorhabdus luminescens strain W14 M13 library"

/dev_stage="primary phase variant"

/note="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janus."

BASE COUNT 166 a 124 c 134 g 211 t 2 others

ORIGIN

Query Match 4.2%; Score 41.8; DB 17; Length 637;

Best Local Similarity 48.9%; Pred. No. 0.26;

Matches 112; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 358 ATCGTCTCACAGCTGTTGAAGACGCTGTCATTGGTACACCAATTGCGAATCGCAACCGAC 417

Db 10 ATCGTTATTCACAGGATGAAGACATCGTTATTTGTTAGTCCATTTGCCAATCGTCACAAATA 69

QY 418 CTGAACCTGGAGGATATCATCGGCTGCTTTGTCAATACGCGAGTGTATCGAATCAACATAG 477

Db 70 TTGATGAACATGTCGCGCTTTGGTTTATTTATCAATTTATTAACCGTTCGGTTTTTCGCTTTA 129

QY 478 ATCATCAGGATACCTTTGGGACTTTGATCAACCAAGTCAAGGTCACGACGACGACGAT 537

Db 130 ACAGGAGGATTAATTTCAATGATGATTTGTTCAAGCGCGCGATATATTCTTGATGCCT 189

QY 538 TCAGAACGAGGATATTTCCGTTTGAGCGCGTGTGATCAGCACTACAGCC 586

Db 190 ATGAATGCTGATTTGCCAATTTGATGAGATTTGTCGATGTGGTACAGCC 238

RESULT 3

LOCUS

AO991640

DEFINITION

Photorhabdus luminescens strain W14 M13 library

Photorhabdus luminescens genomic clone PLG01385F, DNA sequence.

ACCESSION

AO991640

VERSION

AO991640.1 GI:9650309

KEYWORDS

GSS.

SOURCE

Photorhabdus luminescens.

ORGANISM

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

Photorhabdus.

REFERENCE

1 (bases 1 to 797)

AUTHORS

ffrench-Constant, R.H., Waterfield, N., Burland, V., Perna, N.T.,

Daborn, P.J., Bowen, D. and Blattner, F.R.

TITLE

A genomic sample sequence of the entomopathogenic bacterium

Photorhabdus luminescens W14: potential implications for virulence

Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)

JOURNAL

MEDLINE

COMMENT

Contact: ffrench-Constant RH

Department of Biology and Biochemistry

University of Bath

South Building, Bath BA2 7AY, UK

Tel: (44) 1225 826621

Fax: (44) 1225 826779

Email: bssrfce@bath.ac.uk

This is one of a selected subset of flipped clones from the M13

library. For annotation of identified clones (BLASTX, BLASTN and

mapping to E. coli K12 genome) please see ffrench-Constant et al.

2000, Nucleic Acids Res.

Seq primer: M13 Reverse

Class: shotgun.

Location/Qualifiers

1..797

/organism="Photorhabdus luminescens"

/strain="W14"

/db_xref="taxon:29488"

/clone_lib="PLG01385F"

/clone_lib="Photorhabdus luminescens strain W14 M13 library"

/dev_stage="primary phase variant"

1

.....

ORGANISM

Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE

1 (bases 1 to 493)
Smith, T.P.L., Grose, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett,
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,
Perte, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and
Keel, J.W.

TITLE

Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle

JOURNAL
MEDLINE
COMMENT

Genome Res. 11 (4), 626-630 (2001)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@mail.marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTTCCCGTCACGACG

Plate: 66 row: P column: 24

Seq primer: ATTAGTGACACTATAG.

FEATURES

Location/Qualifiers

source

1..493

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="MARC 3BOV"

/tissue_type="pooled"

/lab_host="DH103"

/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."

BASE COUNT
ORIGIN

112 a 106 c 89 g 186 t

Query Match

Best Local Similarity 3.7%; Score 37.2; DB 12; Length 493;

Matches 84; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY

674 GGCTCTCGAGTCGTCACCTGCTAGCAAGCGTACACTCGATTTCACATGGAGTTCCA 733

DB

426 GGCCCTGGAGTCAAAATGAGCAGAGCAAGCGTACAGAGTTTACCAGAGATGCA 367

QY

734 TCTGTTTCAAGAACCGACAGCGCTTAAAGTAGCGTCAACTTTGCCGATGAGCTGTTCAA 793

DB

366 CTGGTCATAGAAACACCCCTTTTCCAACACAGACAACTCTACACATGGACATACCA 307

QY

794 AATGGAGACTGTTGAATTCGTCAGAGTATTTCTTTGAGAT 835

DB

306 GATGGTCACTACTGAAATTGGATTAAATATTCTTTTGAAT 265

RESULT 9

BF643966

LOCUS

NF010A02EC1F1006 Elicited cell culture Medicago truncatula cDNA
clone NF010A02EC 5', mRNA sequence.

ACCESSION

BF643966

VERSION

BF643966.1

KEYWORDS

EST.

SOURCE

barrel medic.

ORGANISM

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE

AUTHORS

1 (bases 1 to 280)
Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.

TITLE

Expressed Sequence Tags from the Samuel Roberts Noble Foundation -
Center for Medicago Genomics Research

JOURNAL

COMMENT

Unpublished (2000)
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7302
Fax: 580 221 7380

Email: radixon@noble.org

Insert Length: 280 Std Error: 0.00

Plate: 010 row: A column: 02

Seq primer: TCACACAGAAACAGCTATGAC.

FEATURES

source

1..280

/organism="Medicago truncatula"

/db_xref="taxon:3880"

/clone_lib="NF010A02EC"

/tissue_type="Cell cultures derived from root tissues"

/dev_stage="Cell suspensions were subcultured every 14

days. Cells were induced six days after subculture"

/note="Vector: Lambda Zap; Cells were induced with yeast

cell wall extracts equivalent to 50ug/ml glucose in the

final concentration. Samples were taken at 0.5, 1, 12 and

24 hours after induction. Equal amounts of RNA from each

time point were pooled and used for mRNA isolation."

74 a 90 c 45 g 71 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

71; Conservative

0; Mismatches

57; Indels

0; Gaps

0;

QY

283

CCCTTCGAGCTTCTGCACGACACACACGACCTTTTCGTCGTTCTTCTAGTCGGT 342

DB

96

CGCGACCAACCTCCCGCAACCAACGACAAATCGCTTCGTCGTTCACTCCCTTTCC 155

QY

343

TCGTCGCGCTCATTCGTCCTCACGCTGTTGAGAGCGCTGTCATTGGTACACCAATTG 402

DB

156

TCGCTTCGCTTACATCTCACCCTACTGTCCTCAAGCTCTCTGTGATACCGCTTTT 215

QY

403

CGAATCGC 410

DB

216

CCAACCCC 223

RESULT 10

BF646516

LOCUS

DEFINITION

BF646516

ACCESSION

BF646516

VERSION

BF646516.1

KEYWORDS

EST.

SOURCE

barrel medic.

ORGANISM

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE

1 (bases 1 to 483)

Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.

TITLE

Expressed Sequence Tags from the Samuel Roberts Noble Foundation -
Center for Medicago Genomics Research

JOURNAL

COMMENT

Unpublished (2000)
Plant Biology Division
The Samuel Roberts Noble Foundation

AW056085
AW056085.1 GI:5928793

Search completed: June 4, 2003, 21:48:21
Job time : 1717.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2003, 17:30:47 ; Search time 2758 Seconds
(without alignments)
10562.693 Million cell updates/sec

Title: US-09-482-788-1_COPY_7000_8000
Perfect score: 1001
Sequence: 1 aacctgttttcacctcg.....tccgtcaaggctgggttg 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	408.6	40.8	10934	8	FSESYNIA	Z18755 Fusarium sc
2	405.4	40.5	2949	8	FSESYNIA	Z48743 Gibberella
3	282.6	28.2	9633	6	BD013055	BD013055 Cyclic de
4	282.6	28.2	9633	23	BD010088	BD010088 Cyclic de
5	240.2	24.0	46899	6	A36768	A36768 Sequence 1
6	240.2	24.0	46899	6	AR050554	AR050554 Sequence
7	240.2	24.0	46899	8	TISACYS	Z28383 T.niveum (A
8	203.2	20.3	1497	8	FPA345016	AJ345016 Fusarium
9	189.2	18.9	1713	6	A36771	A36771 Sequence 4
10	189.2	18.9	1713	6	AR050556	AR050556 Sequence
11	157.2	15.7	3155	8	COPEPSYNH	X96559 C.Oligospor
12	71.6	7.2	10513	1	AF023464	AF023464 Bacillus
13	66.8	6.7	7681	1	AF023465	AF023465 Bacillus
14	66.8	6.7	8371	1	AF087452	AF087452 Bacillus
15	59	5.9	39822	1	BSPEPSYN	Z34883 B.subtilis
c 16	59	5.9	233780	1	BSUB0010	Z99152 Bacillus su
17	58.8	5.9	56917	1	AME16952	Y16952 Amycolatops
c 18	54	5.4	18492	1	AE004666	AE004666 Pseudomon
c 19	53	5.3	22444	1	AE004669	AE004669 Pseudomon
20	51.6	5.2	37200	1	SCF63	AL035640 Streptomy
21	51.2	5.1	107379	1	SHGCP1R	X86780 S.hygrosco
22	51	5.1	8992	1	SV117268	Y17268 Streptomyce
23	50.4	5.0	40989	1	AF204805	AF204805 Nostoc sp
24	50	5.0	41599	1	AB050629	AB050629 Bacillus
25	49.6	5.0	1200	6	AR170980	AR170980 Sequence
26	49.6	5.0	1680	1	AF172065	AF172065 Streptomy
27	49.6	5.0	9389	1	STE250581	AJ250581 Streptomy
28	49.6	5.0	32748	1	AB070951	AB070951 Streptomy
29	49.2	4.9	7178	1	MXA6977	AJ006977 Myxococcu
c 30	49	4.9	10569	1	AE009298	AE009298 Agrobacte
31	49	4.9	31220	1	AE008316	AE008316 Agrobacte
32	48.8	4.9	17377	1	AF237701	AF237701 Pseudomon
33	48.2	4.8	37856	6	AX024319	AX024319 Sequence
34	48.2	4.8	37856	6	AX024212	AX024212 Sequence
35	48.2	4.8	40897	1	AF004835	AF004835 Brevibaci
c 36	48.2	4.8	333500	1	AP003590	AP003590 Nostoc sp
c 37	48	4.8	77457	1	AF210249	AF210249 Streptomy
38	47.2	4.7	26477	1	SCE29	AX433751 Sequence
39	46.8	4.7	6465	6	AX433751	AX433751 Sequence
40	46.8	4.7	8268	6	AX433754	AX433754 Sequence
41	46.8	4.7	14452	1	AB070955	AB070955 Streptomy
42	46.6	4.7	4918	1	AF121767	AF121767 Bacillus
43	46.6	4.7	28798	1	U95370	U95370 Bacillus li
44	46.6	4.7	32386	1	BLAJ5061	AJ005061 Bacillus
45	46.4	4.6	39436	1	ASP269505	AJ269505 Anabaena

ALIGNMENTS

RESULT 1
FSESYNIA
LOCUS 10934 bp DNA linear PLN 11-SEP-2001
DEFINITION Fusarium scirpl esynl gene for enniatin synthetase.
ACCESSION Z18755
VERSION Z18755.3 GI:15591913
KEYWORDS enniatin synthetase; esynl gene.
SOURCE Fusarium equiseti.
ORGANISM Fusarium equiseti
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreales; mitosporic Hypocreales; Fusarium.
REFERENCE 1 (bases 1 to 10934)
AUTHORS Haese,A., Schubert,M., Herrmann,M. and Zocher,R.
TITLE Molecular characterization of the enniatin synthetase gene encoding a multifunctional enzyme catalysing N-methyldeipeptide formation


```
QY 565 GACCCCTTCAGCACTGACAAACCGCTGATGCGCATCGAAAGTCCGCGAGAGGCTTCGGT 624
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8324 GCCCTCTCCAGAAGTCGAGCGCGCTGTCGCGCTCTGCAAGTCCGCGAGAGCTCCAGA 8383
QY 625 CTTTACTTCCATCTGATCATCCATCGAAGATGTTGTTCTGACAGATGCTCTCA 584
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8384 CGCTGGTCCCGCTTACATGTTCTCCGAATATCGTGTGCTGACACGATGCGCTCTCA 8443
QY 685 AGCCCAATGGTAAAGTTGACCGGAAGAACTCTCTCGGAGGCGAAGGTTGTACCGAAGC 744
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8444 ATACTAAGCGCAAGATCGACAGAAGAGCTTACCGGTAGACACGACACTGCCGAGC 8503
QY 745 AGCAGACAGCAGCGCGTTACCGACATTTCCCATCAGTGAGTCSAAGTCATCTTTGCG 804
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8504 AGCAGACTGCGCGCGCTGTCGCGGACTTCCCTATCTCTGATATCGATCAGCTGTGCG 8563
QY 805 RAGAAGCCACTCAGGTGTTGCGCATGAAGGTGACATTACCATCATCTTTCATCTCG 864
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8564 AGGAGCAACTCAGGTCTTTGGAATGAAGGTTGAAATCAGCGATCATTCTTCCAGCTCG 8623
QY 865 GTGGACACTCTCTTGTGGCCAGGAGCTCATTTCTCGTATGACCAAGCACTCAAGGTC 924
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8624 GGGTCACTCTCTCTGCTAGAACTCATTTCTCGCATCCAGCACCGCTCTCCATGTC 8683
QY 925 GTATCACTGTCAAGATGCTTTGACCATCTGTTGATTTGCGGATCTAGCATCTGTATCC 984
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8684 GGGTACTCTGAAGACGCTATTTCGACAGCCCTGCTTTGCCGATCTGGCAGTCATCATCC 8743
QY 985 GTCAAGGCGTGGTTTG 1001
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8744 GTCAAGGACTGCTATG 8760
```

```
RESULT 2
FSENSYN 2949 bp DNA linear PLN 20-FEB-2002
LOCUS Gibberella pulicaris partial esynl gene for enniatin synthetase.
DEFINITION Z48743
ACCESSION Z48743.2 GI:16444855
VERSION enniatin synthetase; esynl gene.
KEYWORDS Gibberella pulicaris.
SOURCE Gibberella pulicaris
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreales; Nectriaceae; Gibberella.
1
Burmeister,J., Haese,A. and Zocher,R.
Highly conserved N-methyltransferases as an integral part of
peptide synthetases
Biochem. Mol. Biol. Int. 37 (2), 201-207 (1995)
96113556
8673002
2
Gliniski,M., Hornbogen,T., Haese,A., Doller,A. and Zocher,R.
Nonribosomal Biosynthesis of N-methylated peptides in fungi
Unpublished
3
Burmeister,J.
Direct Submision
Submitted (17-MAR-1995) Burmeister J., Universitaet Zuerich,
Biochemisches Institut, Winterthurer Str. 190, Zuerich,
Switzerland, CH - 8057
On Oct 25, 2001 this sequence version replaced gi:732698.
Location/Qualifiers
1. .2949
/organism="Gibberella pulicaris"
/strain="BBA 63933"
/isolate="wild type"
/db_xref="taxon:5128"
/clone_lib="pCM2"
/clone_lib="1.5 kb and 2.9 kb PCR fragments in
pBluescript"
1. .2949
gene
```

```
/gene="esynl"
<l..2949
/gene="esynl"
/function="amino acid activation, N-methyltransferase"
/codon_start=1
/product="enniatin synthetase"
/protein_id="CAA8634.2"
/db_xref="GI:16444856"
/db_xref="SPTREMBL:Q00868"
/translaton="EFNTGILKASLAYPLDVRSPVARMKIDILSSVSGNTIVIMGTGV
EDPGGLPQLELVRLTDTFDETIEDVQNMRSPTSILAVVFTSGTSGKPKGVMLTEHR
ALVRLVKSDFNFPSPARMVFNAAEDGASWEMFWLLNGTGYVCIDYLAITLDGKE
LAAYFAKRVNCAFLAPAMLKLYLDAREALKANLDFLAVGGKEFPDRDAEAKNLTVRG
LNIANYGYTEAGMISTCYTIPKDEAFITGVQIRSYNGAYVMDPNQOLAGLVNGE
HLFGVGRTGYTPKELNKNRFIDVTIEGTRAYGTDRMARVGDGLLEFFGRLDNQ
PKMRQTEAGEVESAMLSHKLNAALVIRGGQEGEOLMEVGFIVADDDDDTBEIE
TCNOVEGMDHFEESGMSDSTAVDQSAIGNDFKGMTSMYDGNIDKGMQWMLDDAI
HTLHNGQVPHHVLEITGSGMLFLNPLGLOSQVGLDPSKSAVEFNRAIESSPFAG
KAKVHGNATDVKLGLHPDLVNSVQYFPTPEILTEVIDGLLIAIPSVKRIFLGD
IKSYATNRHFLAARAIHTLNNATKDRVQKQVOELEDREESLVEPAFTTLKERR
PDVVKHEVIEPKMKATNELIARYTAVVHLRDEDEPYVPTKESWIDFEKQMDKK
ALLDLRLSKDAMSVAVSNITYAHTAFERRIVESLDEDSKDDAKDTLGGAAWLSAVRS
ETESRASLTVPLEFELANEGAFRVEVSAAWSONGALDAVFHFPSSNTDRTLQFP
TDNQLRSSLTLANRPLQKLRRAALQVRESLSQSLVPTVMPNIVLVLDMLNTNGK
IDRKELTKRARTLPKQOATAPVDFPISDIEITLCEEATEVEFGMKVEISDHFQLGGH
SLLATKLISRIQIHLRVTVKDVDFSPVADLAVIIHROGLAMQNPVABGQDKQGWSS
RYAPRTEVEKM"
BASE COUNT 701 a 838 c 767 g 643 t
ORIGIN
Query Match 40.5%; Score 405.4; DB 8; Length 2949;
Best Local Similarity 65.3%; Pred. No. 1.4e-112;
Matches 677; Conservative 0; Mismatches 321; Indels 39; Gaps 4;
QY 1 AACCTGCTTTCTTCACTCGTTGAAGACAGGTTTCCAGTCTGCTGGAGCACTGTTGAGA 60
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1838 AGCTTCGATCTTCAGCACTCTGAAGACGAGGACCGGATGTTGTCAAGCACTGTGAG 1897
QY 61 TCCTGCCAAAGAACATGGAAGCTGTGAATGAGTCAAGTCTGATCGATATGCGCTGTTG 120
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1898 TCATCCCCAAGAACATGAAGGCCACCACTCATCGCTACCGCTACACAGCTGTTG 1957
QY 121 TCCAGCTTCGGGTTTCACCTTGGAGATGAGCTTGTGCTTCCGGTTGAGAAGATGACTGGA 180
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1958 TGCATCTGCGGGATGAA---ACGGACGAGCCTGTATACCCCATGAAAAGGATGAGTTGA 2014
QY 181 TCGACTTTCAAGCGAATCAATTGAACGAGAAGTC---ACTGGTGACCTTCTCAAGCTTT 237
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2015 TTGACTTTGAGGAGAAGCAGATGGATAAGAAGCTCTTCTTGATCACTCGCGCTCTCCA 2074
QY 238 CAGATGCTCTATCATGCGCAGTCAGCAAAATTCCTTTCGAAATTCACGGCTTTTGAAGAC 297
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2075 AGGATGCTATGAGCGTGGCTGTTAGCAACATCACCTACGCCACACTGCTTTGAGCGCAG 2134
QY 298 AGGTGCTGCTTCCCTC-----AATAGCAACATCGATG 330
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2135 GCATCGTTGAGTCTCTCGACGAGGATAGCAAGGATGACGCCAAGGATACACTCGCGGTG 2194
QY 331 AGTGGCAGCTATCAACCATTCGTCAGCGCGGAGGCGGACATCATCATCTCGTTCCCG 390
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2195 CAGCATGGCTCTCAGCAGTTCGCTCGAAACGAGAGCGGTGCTCTCACTACTGTTCCCG 2254
QY 391 ACATCTTTTCGCTATGCTGGGGAAGCGGGTTCCGCTGTCAGGTTCAGTCTTTCGACGACAGT 450
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2255 AGCTCTTCGAGATCGCCAACAGGCTGTTTCCGTGTCGAAGTCACGCCGCTCGCCAGT 2314
QY 451 GGTCTCAGAAATGTCATGGAGCTGTTTTCATCA-----TTGTTGTTCCCAAGGCG 504
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2315 GGTCTCAGAAACGCTGCTTGGACGAGTCTTCCACCACCTTCCCATCTCCCAACACGACGAC 2374
QY 505 GTACTCTGTGTAACCTTCTCAGGACCATCACTTCGAGGCTGTGATCTCTCTCACCACATC 564
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2375 GCATCTCATCCAGTTCCCTACGGATTAATCAGCTTCGATCATCACTCACCTCGCCCATC 2434
```


Db	21796	TCAACAAACACGCGAAGATTGACCGCAAGAGCTTGTGGAGAGCTATCGTGGCCCGCA	21855
Qy	742	AGCAGCAGACAGACGCGCTTACCGACATTTCCCATCAGTCAGGTCAAAGTCATTCTTT	801
Db	21856	AGCCAAGGTCAGCGGTAC---TCGGGTAGCCCCCGCAATGAGATCAGAGCTATTCTGA	21912
Qy	802	GCGAAGAAGCCACTGAGGTGTTTGGCATGAAGTTTGACATTACCGATCACCTTCTCAATC	861
Db	21913	GACACGAATTCGAGGACGTGCTCGGAACAGAAGTCAGCGTGTGGATTAACCTTCTTGATC	21972
Qy	862	TCGGTGGACACTCTCTTTGGCCACGAAGCTCATTTCTCGTATCGACCAACACTCAAGG	921
Db	21973	TCGCGCGGCACCTACACTTATGTGCCACGAAGCTCGCGCCCGCTTAGCGCGCTTGATG	22032
Qy	922	TCGGTATCACTGCAAGGATGCTTTTGACCATCTCTGATTTTCGGGATCTAGCATCTGTCA	981
Db	22033	CCCATATTTCCATCAAAGATGCTTTTGATCAGCGGTGCTGCGGATCTTCGCGGTCCA	22092
Qy	982	TCC 984	
Db	22093	TCC 22095	
RESULT 6			
AR050554			
LOCUS	AR050554	46899 bp	DNA linear PAT 29-SEP-1999
DEFINITION	Sequence 1 from patent US 5827706.		
ACCESSION	AR050554		
VERSION	AR050554.1	GI:5973279	
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 46899)		
AUTHORS	Leitner,E., Schneider,E., Schoergendorfer,K. and Weber,G.		
TITLE	Cyclosporin synthetase		
JOURNAL	Patent: US 5827706-A 1 27-OCT-1998;		
FEATURES	Location/Qualifiers		
	1..46899		
	/organism="unknown"		
BASE COUNT	10651 a 13511 c 12510 g 10227 t		
ORIGIN			
Query Match 24.0%; Score 240.2; DB 6; Length 46899;			
Best Local Similarity 55.0%; Pred. No. 7, 5e-62;			
Matches 563; Conservative 0; Mismatches 418; Indels 42; Gaps 3;			
Qy	1	ACCTCGTTTCTTACCTCGTTTGAAGACAGGTTTCCAGGTCTGTGGTGAACATGTTGAGA	60
Db	21076	ACCCTGCATTCTTTACATCTTTGCGTACGCGCTTGGGTGACAAGATCAAGCACGTGAAA	21135
Qy	61	TCCTGCCAAGACATGGAGCTGTGATGAGCTCAGTGCCTATCGATATGCCGCTGTG	120
Db	21136	TTCTGCCAAGACCATGAAGGCTACCAACGAGCTCAGCAAGTACCGATATGCCGAGTAC	21195
Qy	121	TGCACGTTCCGGGTTTCACCTTGAGATGAGCTTGTGCTTCCGGTTTGAGAAAGATCACTGA	180
Db	21196	TACATGTGCTGGCTCGAGAGAACAACTCAACTATACACCAAGTCTCTCCCAACGCCCTGGA	21255
Qy	181	TCGACTTTCAACGGAATCAATTTGAACCAAGAAGTCACTGGTGACCTTCTCAAGTCTTTCAG	240
Db	21256	TAGACTTTCCGCGACAGGCTCTCGACCGCAGACCCCTCATCACTTCTGAAGAGCACAC	21315
Qy	241	ATGCTGCTATCATGGCAGTCAGCAAAATTCCTTTCGAAATCACCGCCTTTTGAAGACAGG	300
Db	21316	AGGATGCCGGGACCGTCGCTATTCGGTAAATATCCCGTACAGCAAGACCACTTGTGAGCGGT	21375
Qy	301	TGCTGCTTCCCTCAATAGCACATCGATGAGTGGCA-----	337
Db	21376	TTGTCAACAAAGTCACTGAGCGAGGATGATATGGAGGAAGGCCAGAACTCACTGAGCGGAT	21435
Qy	338	-----GCTATCAACCACTTCGGTCCAGCGCGGAGCGACTCATCACTATCCGTCGCG	390

Db	21436	CAGCTTGGGTGTCAGCCGTC	CGGATGGCGCTCAAAGCTG	CCCACTCGATGCAATGG	21495
Qy	391	ACATCTTTCCGCAATGTCG	GGGAAGCGGGGTCCGCTG	TCGAGGTCTCTGCACGACAGT	450
Db	21496	ATGTCGAAGAGATTGCTC	AGGAGCGGATACCAAGT	CGAAGTCAGTTGGCGCGTCAAT	21555
Qy	451	GGTCTCAGAAATGGTGAT	TGGACGCTGTTTCCATCAT	TGTTGCTCCCAAGG-----	503
Db	21556	GGTCCAGAAATGGTCGCT	TCGATGCATCTTCCATCA	CTTCCAAACCGCCCAAGAGGGTG	21615
Qy	504	--CGTACTCTGGTCAACT	TTCTCCTACGGACCATCAC	CTTCGAGGGTCTGATCTCTCACCA	561
Db	21616	CTCGCACACTTATTGATT	CCGACGGATTACGAAGC	CCGGAATGTGAACACCTTTAACGA	21675
Qy	562	ATCGACCCCTTCAGCGAC	TGCAAAACCGTTCGTAT	CGCCATCGAAATCCGCGAGAGGCTC	621
Db	21676	ACGTCCTCCCTGAACAG	CAATTCAAGCCGCGCTCT	TGGGACCCAGATCCGCGAGAGGTGC	21735
Qy	622	GGTCTTACTTCCATCGT	ATAGTATCCCATCGAAGAT	CGTTGTTCTGGACAAGATGCCTC	681
Db	21736	AGACCTCTCTGGCGCTT	ACATGATCCCATCGCGAT	CATCGTCTTGATCAGATGCCTG	21795
Qy	682	TCAAACCCCAATGGTAA	GTTGACCGGAAGGAACT	CTCTCGCAGGSCCAAGGTTGTACCGA	741
Db	21796	TCAAACCAACAGCGCA	AGATTGACCCGAAGAGCT	TGTGCGGAGAGCTATGCTGCCCCGA	21855
Qy	742	ACGACGACAGACGAGC	CGCGCTTTACCGCACTTT	CCCATCAGTGGAGTCAAGCTCATTTCTT	801
Db	21856	AGCCAAGTTCAGGGGTAC	---TCGGGTAGCCCCCG	CAATGAGATCGAGGCTATTCTGA	21912
Qy	802	GGCAAGAAGCCACTGAG	GTTGGCATGAAGTTGAC	ATTACCGATCACTTCTTCAATC	861
Db	21913	GAGACGAATTCGAGGAC	GTGCTCGGAACAGAAGT	CAGCGTGTGATGAACCTTCTTTGATC	21972
Qy	862	TCGGTGGACACTCTCT	CTTGGCCACGAAGCTCA	TTTCTCGTATCGACCAACGACTCAAGG	921
Db	21973	TCGGCGGGGCACTACAT	TATGGCCACGAAGCTG	CGCGCCGCGGTATAGCGCGCCTTGATG	22032
Qy	922	TCCGTTATCACTGTCAG	GAAGTGTCTTGACCAT	CTCTGTATTTGCGGATCTAGCATCTGTCA	981
Db	22033	CCCATATTTCCATCAAG	ATGCTTTGATCAGCCCG	GCTGCTGGCGGATCTTCGGCGGTCCA	22092
Qy	982	TCC 984			
Db	22093	TCC 22095			

RESULT 7

TISACYS	46899 bp	DNA	linear	PLN 06-DEC-1994
LOCUS	T.niveum (ATCC34921)	simA gene	for cyclosporine synthetase.	
DEFINITION	Z28383			
ACCESSION	Z28383.1	GI:440168		
VERSION				
KEYWORDS	cyclosporine synthetase; peptide synthetase; synA gene.			
SOURCE	Polypocladium inflatum			
ORGANISM	Polypocladium inflatum			
	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
	Eurotiales; Clavicipitaceae; Clavicipitaceae; Mitosporic Clavicipitaceae;			
	Polypocladium.			
REFERENCE	1 (bases 1 to 46899)			
AUTHORS	Weber, G., Schorgendorfer, K., Schneider-Scherzer, E. and Leitner, E.			
TITLE	The peptide synthetase catalyzing cyclosporine production in			
	Polypocladium niveum is encoded by a giant 45.8-kilobase open			
	reading frame			
JOURNAL	Curr. Genet.	26 (2), 120-125	(1994)	
MEDLINE	95094306			
PUBMED	80011164			
REFERENCE	2 (bases 1 to 46899)			
AUTHORS	Weber, G.			
TITLE	Direct Submission			
JOURNAL	Submitted (29-NOV-1993); Weber G., Biochemie Ges. m.b.H., Molekulare			
	Genetik 2, Kufstein-Schaftenau, Austria, A-6330			

LVQTOAFNAYGPTENTWSTIYPIAEDPFFNGVPIGHAVSNSGAFVMDQNOQTPPGAG
MGLIVTGLGARGYTSSLNTGRF INVDIGEOVRAYRTGDRVRYRPKDLQIEFFGR
IDHQVKTRGRHIEPAEYEAIALSHDLVTDAAVTHVTHSQENDQLEMGVFAVADVRD
ESSNOVOEHTHPSIAIADITTDQOGLGDRPMSWTSMYDGLSIKKSOEWMDDTM
RSLSDOPPGHLEVGTCGMFLNREGGLOSYLEPSPSATAFNKAAKSPFCL
EDRIVEVGTATDIDRLGDHAGLVVNVSAQYFPQSDIYAQLVRLTKVPGERIF
FGMRSHAINRDFIARAVHALGDKATKAELOREVRMESEDELLVDPAAFTSLTTO
VENIKHVEILKRMNRTNELLSSRYAIAHLOVNDLAKPAHKVSPGAWAATKMDRDA
LIRLRTGKTSIDHAIANIPNSKTIVERTICESYDGLGDAKDSNDVLSAARSNA
VKYASIAIDLDIAQAGFVERISCAROWSCALDAVPHHGLSPQSSHHVIDELFT
DHOCRPEALTNP LRAQSRVRERO IRELRQTLTPAYMIPAQIIMWLDKELPLNANGKV
DRKOLTOAQTVPKAKOVAPAPARTIEIUVLCQESDVLGVLDIGENEFDELGGHSL
MAYKLAARISRLRTHSVKREIFDHPRVCDLVLVQOQSAPHDIPVSTKTVGPQPSF
AQGLWFLDQNFATWYLMPLAVRLRGAMNHALTAALALPERRHELLRTTIEYEQG
VGKQNVNVTETIIRIDLSNGDGYLPTLKKEQTAPFHELETERRGWRVLLRLGPGDY
ILSVHHHII SDGWSVDVLFQELGQFYSTAVKGDPLSQTTPL IHYRDFALWOKKPT
QESHEROLOVYEOVLDSAPAEALLTOLPPRSILSGOAGEMSVTIEGALKNLEFCR
VHVTSFVLLAALRAHYRTGSEDATIGTPANRNRPELEQIIGFVNTQIRIVT
NEDETFSVLQVRSATATAFAHQDVFEKIVSTLLPGSDASRNPQLVLMFAVHSQV
NLGELKLENAHSVPEITITREFLEPHFOQDDKLEGSILYSTDEAVSVSSLSV
FOELIRGLNGDPVPISTPLQDGIVDLQROGLDVOKTEYPRDSSVYDFHEQVSN
PDSIALIHGSEKLSYADRESORVARLHRFSFSDTLTAVLAPRSCETIIAFLGIL
KANLAYLPDVKAPAARIDAIVSLPGNKILLCANVTTPKLOEAAIDFVP IRDTFTT
LTGDTLODGP IERPSAQSLAYAMFTSGSRGPKGVAMVHNRVLRVKNVYAKQPA
AARIAHISNAFADASSWEIYAPLLNGGAIYCADYFTTIDFOALQETFOEHEIGAMP
PSSLKQCLVQAPDMISLIDILFAAGDRFSSVDALQAOVLGSGVFNAYGPTENTILST
IYNVAENDSVNGVPIIGSAYSNSGAYIMDKNQQLVPAGVMEVLTDGLARGYMDPK
LDADRFTQLVNSQEVYARTGDRVYRPFQDIEFGRMDQOIKIRGRIEPAEVE
QAFNDGCFQVDAIYITPENQOPMAYFVYAKDNSAREEATTOIEGWEAHEFGGA
YANIEETESALGVDFMGWTSYDGTEDIDKDMREWINDTMRSLDGLKPAGRVLEVT
GTGMINPLNRSQGLERYIGLEPAASAEFVNNAKSPFGLAGRAEVHVGTAADVGLT
QGLTSDMAVINSVAQYPTPEYIAETIKSLVQVPMKRIYGLDMRSWMNRFAAARA
AYSUADNADKVRQNMKEELLYVDPFAFTLASOLODRIOHVELIPKRMKAT
NELSYRYAAVLHISDEPLTYIKIDPEAMINFEGRSLTREALAQVLKENENASVATS
NIPKSVTVRHHIVRSUDODANAPESMDGSDWISAVRTAQQCHTPLASDLFDIAE
DAGFRVSVAROHOSODANAPESMDGSDWISAVRTAQQCHTPLASDLFDIAE
PAQSRRAELLIRGLQTLPLPYMIPSOITILDRMPLNANGKVDRELRARAKTIOKSK
PVEDIYPPRNSVATVCKGFTDLGVVEGTDNFFNLGHSMLMATKLAARLGRQLNTR
ISVRDVEDQVADLAIQIORSAPHEPIKADYTPGPVQPSFAQRLWFLDQNLVNGAT
WYLMPLIRLHLSRLVDALATAISALEORHEPLRTTTEEDGVGVQVOWDRHPKDLRI
IDLSTQPKDAYLAVLHQEOTLFDLATEPCGRVALIRLGEHEHILSVHMHII SDGWS
VEVLDFMRHFYSALRQDPMQIILPLIQYRDLFAWOKTEBOVAHQHOPQLWYFPH
LAOSTPAELLTDLPRPSILSRANELPLTIEGRHDKLRAFRCVHQATPFIYLLAAR
AAHRLTGADATIGPIANRPELNMIGFVNTQCMRIAIEENDFESLVRVARS
TATSANQADVPFESVILLSGSRDASRNPLOVILAVHSQDQLGLTLEGLRDEAV
DSATISREDFEHLFHADRLSGSLYAKELFKLRTIESVSVFLFRLRLDQPLTP
LAVPLTDGVEITASKGLDVPRTDPRDANI VEVFOQHVRATPDALAVKDATSILTY
AQLDQSDRLAIWLSRHHMPETLVGLAPRSCETIIAMFGIMKANLAYLPDINSPA
ARLRSILSAVDGNKVLILGSGSVTAPQOENPEVEAVGIOETILAGTLDKTOGSNARSA
TSLAYVLTSGSGTKPKGVVHRSVTRLAKPSNISKLPQGARVAHLANAFDSTW
EIAITLLNGATVLCDDYHTVDCRTLKEVERESITVWTLMPALLKOCVAEIPETLH
LDLYLTGGDVRGHDAMRARSVKIGMFSGYGTENTVITIEYEVDADEMFGVNGPTIG
KTVNSGAYVMDNRQOVLPSGVVGLVVTGDLARGYTDPLSNKRRYIYITVNGESIR
AYRTGDRVRYRPHDLOIETFGRMDOOVKIRGRHEPEGEVESALLSHNSVODAAVICA
PADQSGAEVFAVAERTEDETOEEEAVDQGWETHETRAYSEVKDLROSEVGN
DFMGWTSNGSEIDKTDHMLNDTMRMLDAREPHGVLEIGTGMWFMFLAKCPG
LQGVYGFPEPSKSAFVNDAAQSPFALKDGRSIVHVGTATDINKAGYIQPRLYVINSV
AQQFTPEYLFYRVAEALYQPSVERIVFGDMRTNAINRDFVSRALHTLEKANKRLV
ROMIYELEANEELLTDPAFTSLRTRLEKIKHVELLPKTRMKNATNELSKYRSYAAVLH
VMSREOSTTHOVSNAWIDFAADLDROTLLINKLKEHKDAGTVAIGNIPYSKTIIVER
FVANKSLSEDMEGONSLOGSAAVAVMAAQCPSILDANDVKEIAQEAQYOVESWA
RQWSONGALLDAIHFFHEPPKEGARTLIEFTDYEGRNVNLTNRPLNSIOSRRLGTQI
REKLQTLPMIPSRIMPLVDQMPNNNKIDRKLVRRAIVAPKPSRASTRVAPRNE
TEAILREDEVLGTEVSDNFDTLGGHLSMATKLAARVSRRLDGAHISIKVDFOQV
LADLAASIQRESAPHERIPQRTYFGAEQFAQRLWFLDQNLGATWYLMPLAIRIR
GOLRVAAALSALFALERHETLRTTEESDGVQVIVGEARNSDRVDHVSITGDGGEY
LEVLREOTVPDLSPEPGWRVCLVKTGEDHVL SVHMHIIYDGHWSVDLLRGLGQF
YSAALRQDPLHMANPLPIQYRDFEAANOREAKQVBEHOROLGYWSQOLVDSFAELLT
DLRPSILNSGRAGSDVTIEGVSYQALQSCFRSRVSTFVVLITVRIAHFRFLTAVDD
ATIGTPIANRPELETIEGVSYNTQCMRISIADDNFEGFVROVRNRTAAYANQDV
PFRIVYALPGVGRSRLSRNPVLQVLMFAQVSDYDQVRLEGLSVMPGEASTRFOME
FHLVPGDKQLTSGVSLDFEQEITQNVDFIQECLRSVLDQPLTPISVLPFSNAIS
NLESILDLLEMTSDYPRDRTVVDLFREQAALCPDSITAVKSSQLTYAQLDEQSDRVA

Query Match 24.08; Score 240.2; DB 8; Length 46899;
Best Local Similarity 55.08; Pred. No. 7.5e-62;
Matches 563; Conservative 0; Mismatches 418; Indels 42; Gaps 3;
Qy 1 AACCTGCTTCTTCCACCTCGTTTGAAGACAGAGTTTCCAGGTCTCGTGGGAACATGTTGAGA 60
Db 21076 ACCCTGCAATCTTTTACATCTTTTGGCTACGGCTTGGGTGAGAGATCAAGACGCTCGAAA 21135
Qy 61 TCCTGCCAAAGAACATCGAAGCTGTGAATGAGCTCAGTGGCTATCGATATCCCGCTGTG 120
Db 21136 TTCTCCCAAGACCATCAAGGCTACCAACAGAGCTCAGCAAGTACCGGATATCCCGCAGTAC 21195
Qy 121 TGCACGTTCCGGGGTTCACCTTGGAGATGAGCTTGTGCTTCCGGTTGAGAAAGATGACTGGA 180
Db 21196 TACATGTGCGTGGCTCGAGAGAACAACTAACATATACACCAAGCTCTCCCAACGCTGGA 21255
Qy 181 TCGACTTTCAAGCGCAATCAATTGNAACAGAGAGTCACTGGGTGACCTTCTCAAGTCTTCAG 240
Db 21256 TAGACTTTGGCGGACAGCGGTCTCGACCGGACGACCTTCATCACTTCTGGAAGGACACA 21315
Qy 241 ATGCTGTATCATGGCAGTCAGCAAAAATTCCTTTTCAAAATCACGGCTTTTGAAGACAGG 300
Db 21316 AGGATCGCGGGACCGCTCGCTATCGGTAATATCCCGTACAGCAAGACCATGTTGAGCGGT 21375
Qy 301 TCGTTCGCTTCCCTCAATGAACAAATCGATCGATGAGTGGCA----- 337
Db 21376 TTGTCAACAAGTCACTCAGCGAGGATCATATGAGAGGAAGGACCAACTCACTGACCGAT 21435
Qy 338 -----GCTATCAACCATTCGGTCCAGCGCGGAGGCGGACTCATCACTATCCGTTCCCG 390
Db 21436 CAGCTTGGGTTGACCGCGTCCGATGCGCGCTCAAACTGCGCCATCGATCAATG 21495
Qy 391 ACATCTTTCGCAATTCGTTGGGGAAGCGCGGTTCCGTGTTCGAGGTTCAGTTCGACGACAGT 450
Db 21496 ATGTCAGGAGATGTCCTCAGAGCGGGATACCAAGTTCGAGTTCAGTTCGGCGGCTCAAT 21555
Qy 451 GGTCTCAGATGTCGATTCGAGCGCTGTTTTCACCATTAFTGTTGCTCCCAAGG----- 503
Db 21556 GGTCCCAAGAATGGTGGCTCGATGCCATCTTCCATCACTTCCGAACCGCCCAAGGAGGTG 21615
Qy 504 --CGTACTCTGCTCAACTTTTCTTCCACGACCATCACTTCGAGGGGTCTGATCTCCTCACC 561

DEFINITION	Sequence 4 from patent US 5827706.		
ACCESSION	AR050556.		
VERSION	AR050556.1	GI:5973281	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 1713)		
TITLE	Leitner,E., Schneider,E., Schoergendorfer,K. and Weber,G.		
JOURNAL	Cyclosporin synthetase		
FEATURES	Patent: US 5827706-A 4 27-OCT-1998;		
source	Location/Qualifiers		
	1. .1713		
	/organism="unknown"		
BASE COUNT	389 a	491 c	469 g
ORIGIN			364 t

	Query Match	18.9%	Score 189.2	DB 6	Length 1713
	Best Local Similarity	59.5%	Pred. No. 2.3e-46		
	Matches 396	Conservative 0	Mismatches 258	Indels 12	Gaps 4
Qy	338	GCTATCAACCATTCGGTCCAGCCGAGGGCGACTCATCACTATATCCGTTCCCGACATCTT	397		
Db	992	GATATCGGCACCTCAATACGGCGGAAGGAATGCCCTGCTCTCTCAGTGCGCCACCTGAT	1051		
Qy	398	TCCGATTCGTGGGAAGCCGGTTCGCTGTCGAGGTCAAGTCTTCGCACGACAGTGGTCTCA	457		
Db	1052	TGAGATTGGTAAAGGGATCGGCTTCCAAGTTTGAGACGACGCTGGGCTCGACAACACTCCCA	1111		
Qy	458	GAATGTGCAATTGGAGCGCTCTTTTCCATCATCTTGTCTCCAA-----GGCGGTAC	508		
Db	1112	GCGGCGGGAATCGATGCTGTTTTCCACCAGATTCGAACAACCAAGACACTCGGGTCATGT	1171		
Qy	509	TCGTGTCAACTTTCCTACGGAACCATCACTTCGAGGGTCTGATCTC-CTCAACAATCGAC	567		
Db	1172	CATGTTTCAGGTTCCCAACTGAACACAGGGGCGGTCTTCGAGCAGTCTCAACAATCGCC	1231		
Qy	568	CCCTTCAGCGACTGCAAAACCGTCGATGCCATCGAAGTCCGCGAGAGGCTTCGGTCCCT	627		
Db	1232	CGTACACCTGGTTCAGAGCGCGCGGTGGAGCAAGGTCCGCGAGCGCTGCAATCGC	1291		
Qy	628	TACTTCCATCGTACATGATCCCATCGAATCATCTTCTTGGACAAGATGCCTCTCAACG	687		
Db	1292	TGCTTCCATCGTACATGATTCCTCTCGGATCATGTGCTCGATCAGATGCCTCTCACGT	1351		
Qy	688	CGAATGGTAAAGTTGACCGAAGGAACACTCTCTCGAGGGCAAGGTTGTACCGA-AGCAG	746		
Db	1352	CCAACGGCAAGTGGATCGCAAGAAGCTCGCTCGACAAGCCGGGTCTATCCCAACAATTG	1411		
Qy	747	CACACAGACAGCCGCTTACCGACATTTCCCATCAGTGAGTCCGAAGTCT-ATTCTTTCCGA	805		
Db	1412	CCGCAAGCAGTTGGACTTTGTGGCGCGACGCGAGGAATCAGGTCGGTCTCTGCGCA	1471		
Qy	806	AGAAGCACTAGAGTGTTTGGCATGAAGTTGACATTACCGATCACTTCTTCAATCTCGG	865		
Db	1472	AGAAATTACCGATCTACTAGGCGTCAAGTTCGGCATTCAGACAACTTCTTCGAGTTGG	1531		
Qy	866	TGGACACTCTCTTTCGGCAACAGCTCATTTCTCGTATCGACCACAGACTCAAGTCCG	925		
Db	1532	CGGCCAATTCGCTGCGCCACGAACACTGACGCGCACGTCCTAAGTCGCAGACTGGACGCCG	1591		
Qy	926	TATCATCTCAGAGTCTTTGACCATCCTGTATTTGCGGATCTAGCATCTGTCAATCG	985		
Db	1592	TGTCATCTGTAAGCAGATCTTTGACCAGCCAGTACTGTCTGCTTGCCTTCTATTTCG	1651		
Qy	986	TCAAGG 991			
Db	1652	TCAAGG 1657			

RESULT 11
COPEPSYNH
LOCUS
COPEPSYNH
3155 bp
DNA
linear
DIN 03-FEB-1998

```
DEFINITION C.oligospermum gene encoding synthetase.
ACCESSION X96559
VERSION X96559.1 GI:1770179
KEYWORDS peptide synthetase; peptidase synthetase.
SOURCE Cyndrotrichum oligospermum.
ORGANISM Cyndrotrichum oligospermum
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Chaetosphaeriaceae; mitosporic Chaetosphaeriaceae;
Cyndrotrichum.
REFERENCE 1 (bases 1 to 3155)
AUTHORS Bernhard,F.
TITLE Identification of genes encoding for peptide synthetases from
gram-negative bacteria and filamentous fungi
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3155)
AUTHORS Bernhard,F.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-1996) F. Bernhard, Freie Universitaet Berlin,
Institute of Crystallography, Takustr. 6, 14195 Berlin, FRG
FEATURES
    Location/Qualifiers
        1..3155
            /organism="Cyndrotrichum oligospermum"
            /strain="NRR1 18230"
            /db_xref="taxon:72418"
            /clone_lib="pCB11"
            <1..>3155.
            /function="amino acid adenylate forming module"
            /note="SD2 214-103"
            /codon_start=2
            /product="peptidase synthetase"
            /protein_id="CAA65395.1"
            /db_xref="GI:1770180"
            /db_xref="SptREMBL:P97961"
            /translation="DPQOLVPVGMGLVVTGDLRGVYTDPLDQGRFVQIRIKGE
SVRAYRGDRARYRPGKQIEFFGRMDQIQIRGRHRELAEREVHMLHDAVRVAVV
LRQSDQPEMIGVVRADETVQDLSRRAGVAVNSANWEEQFEIQTEKERLNLQNL
EEFSDVGEVGVADSFDDGLGSHLMATKLAARISRRLNARVSVKEVDFQVPLADLAI
VIQRGSPHPNLTAPYSGPVQSFAGQRLWFLQINLGASWYLMPLAARVLRPLPTE
ALTATLALQORHEILRTTFRDQGVGVVREHLEKRLVIDVPADLDGGLQPLQKQ
BQTSFDLASPGWRVSLIRANDHVLISVMHHIISDQWSIDLIRBELGOLYSAALR
GYDQSQSLPILQYRDFSVQKQAEQVAERQLEYWTTTLRADDSPAELLTDLPRPT
VLSGNVQVLAIDGSLXERKURAFRCATQYTSFVLLAAFRATHYRJTGAEDATIGTP
IANNRPELNLGIEFFVNTQCMRTVDRTDFTLMOQVRSTTTAAAFANQDPERIV
SALLPGSDTNRNPLVQLVHSHQKDLGMLGVGEVGPVPTAAGTFDEFLHFLQE
VRLGGTVLFDLFPDPTIRGMVTFQEVHLHVDQPHPTITSLPLANGLAGLNCMG
LLDIERSDYPRESLVDVFEQVACPDATAVIDSSRLTYTOLDROSDEVAGHLRR
HMAETLVGLAPRSCQTIYAFGLKANLAYLPDWNVPAARTEAILLSAVEGHKLVL
LGSDVAAPKRVQDVMVRVDDTLGYDILDYANSVAVRPSPTSLEYVMFTSGSTGKPK
GVMEHRAVLRLVNSVNSILPMTPRVHLNLGFDISVQEVYTAALNGTGLWCIDY
FTTLDNSALETVFMQEQVRVAMLTAPALLKLCGLADIPSVLGLDLVLFNVGDRFDSRDAL
AAQALIRGVVNAVYGTENG1"
BASE COUNT 683 a 914 c 859 g 699 t
ORIGIN
Query Match 15.7%; Score 157.2; DB 8; Length 3155;
Best Local Similarity 61.9%; Pred. No. 1.6e-36;
Matches 249; Conservative 0; Mismatches 153; Indels 0; Gaps 0;
QY 590 TCGTATCCCATCGAAGTCGCGAGAGCTTCGGTTCCTTACATTCATCGTACATGATCCC 649
DB 442 TCAGACTGAAAAGAGATCCGTAATCGGTTGCAGAACTCGTATCGTACATGATGTC 501
QY 650 ATCGAACATCGTGTCTGGACAGATCGCTCTCAACGCCAATGTAAGTTGACCGGAA 709
DB 502 GGCTCGGATCACAATATAGACCATGCCCTTAAATGCAATGCAAGGTTGACCGGAA 561
QY 710 GGAATCTCTCGAGGCGAAAGTTGTACCGAAGCAGCAGCAGCAGCGCGCTTACCGAC 769
DB 562 AGACCTTCTCGAAGGGCTCAGACCGTCTCGAAGCTCAGAAGTTGCCATCTGCACGCGT 621
QY 770 ATTTCCCATCTAGTGAAGTCGAAGTCATCTTTGCGAAGAACCCATGAGGTGTTGGCAT 829
```


Y	AGRPHADLEAVLGMFVNTLALRTHPAGNKTFTDLEKVKQTALEAYEHODYPEEEL	Y	QDLGVQREMSRDLFTTLLVQNMKEQTLTKMLNDLRLWMDLQHPISKDFLSLVYTHE
Y	AEIYQVQETALFEEDTVQWAGLFTTLVNTVSSPDMELDQIPILEEFERSFDLIS	Y	AEIYQVQETALFEEDTVQWAGLFTTLVNTVSSPDMELDQIPILEEFERSFDLIS
Y	COLEETGYRLKOTLHLLEBQQAQKPDQPAVVEDEIITYKELNLEKNDSDILLDR	Y	COLEETGYRLKOTLHLLEBQQAQKPDQPAVVEDEIITYKELNLEKNDSDILLDR
Y	GVAAETVYALMGRSRPMLJGLVAILKAGAGYIPIDPDIPEERIRLLKNDSDILLDR	Y	GVAAETVYALMGRSRPMLJGLVAILKAGAGYIPIDPDIPEERIRLLKNDSDILLDR
Y	QSDNLDPADFGEIVLHLSQNLGKRLSNPNVVPDLSLAYMYITVSGSTRGPKGVQV	Y	QSDNLDPADFGEIVLHLSQNLGKRLSNPNVVPDLSLAYMYITVSGSTRGPKGVQV
Y	EHOSANFLNSLOFRPLNSDVLHKTYSFSDASIWELFWPFGASVYQGLQPGGK	Y	EHOSANFLNSLOFRPLNSDVLHKTYSFSDASIWELFWPFGASVYQGLQPGGK
Y	EPDMILVIEEQQITAMHFVPSMLHAFPLNGVPIKTNRLKRVFSGGEQGLAHVL	Y	EPDMILVIEEQQITAMHFVPSMLHAFPLNGVPIKTNRLKRVFSGGEQGLAHVL
Y	SRFCELPDVTLLNSYQGPATVEAFAFDCLDEKLDRIPIGPIHVRVLIINQOK	Y	SRFCELPDVTLLNSYQGPATVEAFAFDCLDEKLDRIPIGPIHVRVLIINQOK
Y	MLPAGCIGELIYAGVARGVYLNRPETLRFDDPFGYPMRYMTKDLARMLPDGQV	Y	MLPAGCIGELIYAGVARGVYLNRPETLRFDDPFGYPMRYMTKDLARMLPDGQV
Y	EFLGRLDQKQIRGYITEPEIEAALRSIEGVREAAVTVTESEGAELCAYAGELGRN	Y	EFLGRLDQKQIRGYITEPEIEAALRSIEGVREAAVTVTESEGAELCAYAGELGRN
Y	EVKQLETLPLGVMYPAHIITEMQVWPSPGDKRKALPADPADGAADRETYTPRNLT	Y	EVKQLETLPLGVMYPAHIITEMQVWPSPGDKRKALPADPADGAADRETYTPRNLT
Y	MKLSOLWEEVLKSGPCHIDNFDGRGSHLKATVLSRIAKFCVQVPLKDVFAHTV	Y	MKLSOLWEEVLKSGPCHIDNFDGRGSHLKATVLSRIAKFCVQVPLKDVFAHTV
Y	EGLASVISTESPSEYAIKPAKRGKRYVSSASAKRMVYLOQLEDGGTGYNMPVLELE	Y	EGLASVISTESPSEYAIKPAKRGKRYVSSASAKRMVYLOQLEDGGTGYNMPVLELE
Y	GKLDITMDRAVFKKLRHLSRLSTFTTDDGEPVORIHDEVPFRHTSVLGGOTBOE	Y	GKLDITMDRAVFKKLRHLSRLSTFTTDDGEPVORIHDEVPFRHTSVLGGOTBOE
Y	AASAFIOPDLSQAPFVQVYVSDGRHLLVDMHHIISDGVSANVLIRFEGELVAN	Y	AASAFIOPDLSQAPFVQVYVSDGRHLLVDMHHIISDGVSANVLIRFEGELVAN
Y	QELPAPHQIKDYAVQWQAFKVGQYTKYQAYVLOKLEGPLVDLPADYARVPVRSF	Y	QELPAPHQIKDYAVQWQAFKVGQYTKYQAYVLOKLEGPLVDLPADYARVPVRSF
Y	AGDQSLPTDQELTAGLHKLARNGSTLVYMLVLAAYTALLARVSGQEDIIVGSPVAGR	Y	AGDQSLPTDQELTAGLHKLARNGSTLVYMLVLAAYTALLARVSGQEDIIVGSPVAGR
Y	PHKDFLPLIGMVFNTLALRARPEGKPPAFQLOEVRETALEAYEHODYPEEELVDLKG	Y	PHKDFLPLIGMVFNTLALRARPEGKPPAFQLOEVRETALEAYEHODYPEEELVDLKG
Y	VTRDSRNPVDEAMLVQNTFOOTFELADLIQLOPAGVSHSHAKFDITLQASGDEQIH	Y	VTRDSRNPVDEAMLVQNTFOOTFELADLIQLOPAGVSHSHAKFDITLQASGDEQIH
Y	LKFEYSIATLQEAETIKQWYTHLTKLUSHVKNRNTLNLDLILTHEERNKLNFNSGG	Y	LKFEYSIATLQEAETIKQWYTHLTKLUSHVKNRNTLNLDLILTHEERNKLNFNSGG
Y	QENEHVQITIDQEFKQALPPNALPVSRENTLYTOELDEWSNGTALWARLDGRVAPD	Y	QENEHVQITIDQEFKQALPPNALPVSRENTLYTOELDEWSNGTALWARLDGRVAPD
Y	TPVGIIMPFSEFMAIALVGMWAGGCVPIDEYPAERKRYITIGDSRTKILLITLPADA	Y	TPVGIIMPFSEFMAIALVGMWAGGCVPIDEYPAERKRYITIGDSRTKILLITLPADA
Y	QEALEDPEGEIFAENATYSPQVSPQPHLAYIITYSTGTTGQPKGVWVSHKGTIA	Y	QEALEDPEGEIFAENATYSPQVSPQPHLAYIITYSTGTTGQPKGVWVSHKGTIA
Y	NTLQWRNRNAVTLNPTVLOLQFSFGDFTLSMTPLLSGATVLPDENAKDILSIK	Y	NTLQWRNRNAVTLNPTVLOLQFSFGDFTLSMTPLLSGATVLPDENAKDILSIK
Y	HDLSHRIYTHMILVPLVIRALLOYIOPEDAKTLRVVTLAGEAADRELDORSMAICNP	Y	HDLSHRIYTHMILVPLVIRALLOYIOPEDAKTLRVVTLAGEAADRELDORSMAICNP
Y	ELANECPTNSVATKTHMHWKQKNGVIGRPIDNTIEVLLINGDQLOPQVAGELCITA	Y	ELANECPTNSVATKTHMHWKQKNGVIGRPIDNTIEVLLINGDQLOPQVAGELCITA
Y	GTGLAGVYNLPDLTAKTFRVQHPYQEKRMRTGDAARMWADGTEIYLGMDQDVVKIR	Y	GTGLAGVYNLPDLTAKTFRVQHPYQEKRMRTGDAARMWADGTEIYLGMDQDVVKIR
Y	GHRTVETKIESVIRRSISGVKEAVVLAARETAQTELCAYIVAEQDENTEMLRALNKQ	Y	GHRTVETKIESVIRRSISGVKEAVVLAARETAQTELCAYIVAEQDENTEMLRALNKQ
Y	LPLYMLQFTLDSLPSPNGKLDRLRAPFPANQVTRFAPHTSMEQKLADIWGEV	Y	LPLYMLQFTLDSLPSPNGKLDRLRAPFPANQVTRFAPHTSMEQKLADIWGEV
Y	LGAEKIGIDDSFELGDSITKALQVLSARLHRKQMSKDLFSHSTIKELAAVIRDS	Y	LGAEKIGIDDSFELGDSITKALQVLSARLHRKQMSKDLFSHSTIKELAAVIRDS
Y	TSASQAPVEGEQVSPQVKWFLSODIKEKHFNQSVMLHRSTSIQDALUKLKAITDI	Y	TSASQAPVEGEQVSPQVKWFLSODIKEKHFNQSVMLHRSTSIQDALUKLKAITDI
Y	HDRLARMTVQSEQGVQVYNNRVPVSHSDALYSLOILNLKGSEVNEVESILKIHVRDI	Y	HDRLARMTVQSEQGVQVYNNRVPVSHSDALYSLOILNLKGSEVNEVESILKIHVRDI
Y	QONINLEKPLLAGLEFRTDDGDFELTAHHFIVDGVSHRVLLEDLATGYRQAVNGED	Y	QONINLEKPLLAGLEFRTDDGDFELTAHHFIVDGVSHRVLLEDLATGYRQAVNGED
Y	RDVPSLNTKETIEALLKHNYSYNTDQDMLLASVILARLWTKQSAVKLSLESHGR	Y	RDVPSLNTKETIEALLKHNYSYNTDQDMLLASVILARLWTKQSAVKLSLESHGR
Y	EVDLGDIVSRVGVMTAIPVFLKDSLDLTGKDQBEYLVHVLKTKTDLRCVPRDGF	Y	EVDLGDIVSRVGVMTAIPVFLKDSLDLTGKDQBEYLVHVLKTKTDLRCVPRDGF
Y	GYGLVLYPPERTGYTADISFNYLQFENSRTEYATIDNTIDENTESFSPJGDEIS	Y	GYGLVLYPPERTGYTADISFNYLQFENSRTEYATIDNTIDENTESFSPJGDEIS
Y	TSNWRQITLDISAMAAEGKMTVMNVETSRFQHTKIAQLCENCHYLLKLLIDHCLNKT	Y	TSNWRQITLDISAMAAEGKMTVMNVETSRFQHTKIAQLCENCHYLLKLLIDHCLNKT
Y	ETEBKTTSDFDHLETDALQEIADMLSFH"	Y	ETEBKTTSDFDHLETDALQEIADMLSFH"
BASE COUNT	3016 a 2429 c 2550 g 2518 t		
ORIGIN			
Query Match	7.2%; Score 71.6; DB 1; Length 10513;		
Best Local Similarity	52.8%; Pred. No. 3e-10;		
Matches	208; Conservative 0; Mismatches 174; Indels 12; Gaps 2;		
QY	603 GAAGTCGCGAGAGGCTTCGGCTCTTACTTCCATCGTACATATCCCATCGAACATCGTT	QY	603 GAAGTCGCGAGAGGCTTCGGCTCTTACTTCCATCGTACATATCCCATCGAACATCGTT
DB		DB	
QY	5552 GAAGTCGCGAAGCAGCTTGACACTTTGCITCCGGGCTACATGATCCTGCTATATCAT	QY	5552 GAAGTCGCGAAGCAGCTTGACACTTTGCITCCGGGCTACATGATCCTGCTATATCAT
DB		DB	
QY	663 GTTCTGGCAAGATGCTCTTCAACGCCAATGGTAAAGTTGACCGGAAGGAACCTCTCTCGC	QY	663 GTTCTGGCAAGATGCTCTTCAACGCCAATGGTAAAGTTGACCGGAAGGAACCTCTCTCGC
DB			

Db 5900 GAAGGTTTGCGTCTGTCTATCAGCAGGAAGAACGG 5933

RESULT 13
AF023465

LOCUS AF023465 7681 bp DNA linear BCT 14-OCT-1997
DEFINITION Bacillus subtilis fengycinn synthetase FenE (fene) gene, complete cds.

ACCESSION AF023465
VERSION AF023465.1 GI:2522213

KEYWORDS
SOURCE Bacillus subtilis.
ORGANISM Bacillus subtilis

REFERENCE 1 (bases 1 to 7681)
AUTHORS Chen, I., Lin, G., Shu, H. and Liu, S.
TITLE Analysis of the Fengycinn Synthetase Gene fene
JOURNAL Thesis (1997) Microbiology and Immunology, Chang-Gung University
REFERENCE 2 (bases 1 to 7681)
AUTHORS Chen, I.
TITLE Direct Submission
JOURNAL Submitted (06-SEP-1997) Microbiology and Immunology, Chang-Gung University, 259 Wen-Hua 1st Road, Kwei-San, Taoyuan 333, Taiwan, R.O.C

FEATURES
source Location/Qualifiers
1. .7681
/organism="Bacillus subtilis"
/strain="F29-3"
/db_xref="taxon:1423"
5. .7681
/gene="fene"
59
/gene="fene"
17. .7681
/codon_start=1
/gene="fene"
/transl_table=11
/product="fengycinn synthetase FenE"
/protein_id="AAB80956.1"
/db_xref="GI:2522214"
/translation="MPQPEIODIYPLSPWQGMFLHSLYDFQSRAYPEQASFTINGQ
LDDRFQKMDAVDRKIEFLFRITFYKNVAKPRQVYLVKNRSHHHVFDLGMGLSILNREKE
HCTEAPKQDKKDFDQSLMIRISLVKAPERYVCITWSHHHFDLGMGLSILNREKE
LHLYQALGGRPLTPQVPGYGYTKLMOQDREAAAYWKKRLQHFKEKASLPKRNE
QVSEGLQWPTTISEKTSYDQKTAASGATLTVQALMGILQLKFNCRCDVAFVGS
VWSGSELDVENIVGLFINIPTRVQSKALSFGLSVRMOKMETEABAYSYFFLYD
IQAASKQELDHIIVFENVPVTOEIGLNOTGSGFSEVFNMAEETIYNVCSYKVI
PGETLVRIINFTGLYEPGIIAEIKDYLOHMISAVISDPSPVSKMTLLDHDKTKQMV
AENTGVSALATPHLGFERQVATLDDPQVAVLPDPAYPERLSYQALVTLNRCARLRA
HGTIKESIYGVLSERSPQVLPDPAVLAAGVLPDPAYPERLSYQALVTLNRCARLRA
TOPGLKAPYSKTLVEPLDANLASECHSLQADSDSLAYIVITSGSTGRPKGVAV
EHQAVSYFTGMQFPLEDDIIMKMTSFSDASVQMLFWMTLSGASYLILPSPGWEK
DPAIMVKALRECVYTAHPIPALMNSFLDQAEAPRSLKRVFAGGELPAPHAARFA
SILPTSLVHGVPTEADVAAYACDPELDKRLRLPIGKVPVGCARLYVLDPLHVAO
PVGVAGELYIAGAVYGVLYNRPELTTEERFLDDPPYGERMYKTDGLARWLPGDQVEP
LGRLDQVQKIRIPIEGBIEAALRSIEGVEAAVTVRTESGEALCAEYAGSGRNEV
RKOLEFTLPGYIMPAHIEEMQVPTPSGKLDKRALPAPDGAADRETVTAPRNLTEV
LSQWBEVLKSGSPVGIHDMFDRGGHSLKATLVASRIAKEFGVQVPLKDPVFAHPTVEG
LASVISEGTESPYEAIKPAGKRETVPVSSAKRMVYVLOLEDGGTGYNMPVLELEGK
LDLTRMDAVFKLILKRHESLRVTSFTDTPDGPVQRTHDEVPRFHTSVLGGOTEQEA
SAPLTPDHLDSQAPLPRAGIVKYSDGRHLLLVDMHHILSDGVSVNTLIRFEGELYANQE
LPAHLQYKDIYVQWQAFKGETYKQRAYWLKQLEGELPVLDDPADTARPPVPTFAG
DQVSEFTLDQELTAGLHKLARENGSTLYWVLLAAYTALLARISQEDDIIVGSPDIAGRPH
KQELPLTGHFVATLALRAPEGGKPPFAQLQEVRETALPAYEQDYPEELVDKLGMT
RDMSPNPLDGMVFMVLQNMDSMQLKDLCRPNATVEHQVSKFDLTLYAYEANGSM
FOMEYSYDLYQKNTMELWLYLNMMLRALIODEAALGTIHLVDENETHFLTHELNC
KREYPRHETISRLFELOQAOTPDALAVAGDEQTLTYKELNIRANRIAAVLRKGVGPE
TVYALLTTPTEIAGMGLIKAGNAYLPADKLPADRLAISYMLSDSGAKILLQSEKAN
NORLDLEKCEIIVGIEIQOQETKFNFAKDLADRLISYMLSDSGAKILLQSEKAN
VIRLVKNSNYIEFTPEDRLSLTSSLGDFVVTVEITGLPLNGAALTKDKETFLDSHQL
KRYIEQNGTITWMLTASLHNLTSEKRTFSRLNLIIGGEALSASHVNRILKQACPEL
SWNGQVGPENTTFTSCPIETKILYSPYIIGPRIGNSTAFILPNKMGMLQPIGAGVELC
SWNGQVARYLGRPDITKSEYVNPFPAPGEOMYRTGDLARWLPLDITIEVGRVDDOAV

Db	3530	-----GCTCTGACGGAGCGCAGATCGGGAACCACTACACAGCACCAGAAATCTGA	358
Qy	783	GAGGTCGAAGTCATTTCTTTGCGAAGAACCCACTGAGGTGTTTGGCATGAAGGTTGACATT	842
Db	3582	CGAGATGAAGCTTCCACGCTGTG-CGAGAAGTCTGAAAAGCGCCCTGTGCGCAT	3640
Qy	843	ACGATCATCTTCTCAATCTCGGTGGACACTCTCTCTTGGCCACGACGCTCATTTCTCGT	902
Db	3641	CAGCACAACATTTTTTGACCGGGCGGCAATCTCTTAAAGCCACCGCTTGTCTCCCG	3700
Qy	903	ATGACCAACGACTCAAGTCCGTATCACTGTCAAGATGCTTTGACCATCTCTGATTT	962
Db	3701	ATGCCCAAGGAATTCGCGTCCAAAGTCCGCTGAAGGATGTGTGCGCCATCCGACGGTA	3760
Qy	963	CGCGATCTAGCATCTGTCAATCCGTCGAAGGCGTGG	996
Db	3761	GAGGTTTGGCTCTGTCTATCAGGAAGNAACGG	3794
RESULT 15			
BSPEPSYN		39822 bp	DNA
LOCUS			linear
DEFINITION			BCT 29-JAN-1997
			B. subtilis genes for peptide synthetase and penicillin binding
			protein.
ACCESSION			234883
VERSION			234883.1
KEYWORDS			GI:1805667
SOURCE			penicillin binding protein; peptide-synthetase; pps operon.
ORGANISM			Bacillus subtilis.
REFERENCE			Bacillus subtilis.
AUTHORS			Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
			1 (bases 1 to 39822)
			Tognoni,A., Franchi,E., Magistrelli,C., Colombo,E., Cosmina,P. and
			Grandi,G.
TITLE			A putative new peptide synthase operon in Bacillus subtilis:
JOURNAL			partial characterization
MEDLINE			Microbiology 141 (Pt 3), 645-648 (1995)
PUBMED			95227362
REFERENCE			7711903
AUTHORS			2 (bases 1 to 39822)
JOURNAL			Grandi,G.
			Direct Submission
			Submitted (27-JUN-1994) Grandi G., Enrikerche S.P.A., Genetic
			Engineering and Microbiology, Via F. Maritano,26, S. Donato
			Milanese (MI), ITALY, 20097
			Revised by [3]
REMARK			3 (bases 1 to 39822)
AUTHORS			de Ferra,F. and Tognoni,A.
TITLE			Direct Submission
JOURNAL			Submitted (24-JAN-1997) F. de Ferra, Enrikerche S.P.A.,
			Environmental Biotechnology Laboratories, Via F. Maritano,26, S. .
			Donato Milanese (MI), ITALY, 20097
			Revised by author 27-JAN-1997
REMARK			On Jan 29, 1997 this sequence version replaced gi:509465.
COMMENT			Location/Qualifiers
FEATURES			1..39822
source			/organism="Bacillus subtilis"
			/strain="168"
			/db_xref="taxon:1423"
			<1..129
CDS			/codon_start=1
			/transl_table=11
			/label=CDS1
			/product="hypothetical protein"
			/protein_id="CAA84365.1"
			/db_xref="GI:509466"
			/db_xref="SWISS-PROT:P39840"
			/translation="PYTWTNAVNDLPSSLTGLQVLEPGEETAKSSITIELNHQ"
			147..156
RBS			163..1638
gene			/gene="bbp"
			163..1638
CDS			/codon_start=1

RBS 9717...9725
gene 9731...17413
CDS 9731...17413
/gene="pps2"
/codon_start=1
/transl_table=11
/label=CDS4
/product="peptide synthetase ORF2"
/protein_id="CAA84361.1"
/db_xref="GI:509469"
/db_xref="SWISS-PROT:P39846"
/translation="MAQSAQODIYPLSHMQEGLMFLSHLMDFFSKKAYIEQTSFTITGN
LCVDFQKSLMLLSRYDIFRTIFIKVEPDLTGPOQVYLSNRELTYVEDISRLQK
QOTLIDAMFTKDRKQDLQKPLMLALDFKQDSOYTCVMTTHHIIIMDGWLCGLIILK
EFFSMYSLKNSPVOLGTPVPSRYIEMIGEODEETAAYSEYLYKYGNTASTPRI
KRRADGNYKADQVSFLAPDMVEKLTEAQNWSGYTLNLFMSWGLVLLHRYNAADDA
VFGVISORGAIDGIESMVGFLINTVPVRSAGGITFSSLVKAVQEDILSSEOHGY
YPLVEIQHNSPLKQGLDHIIFVFNYPVOLHQALSVESENDEGALKLSDISMSQNTY
DENIVIPGESFYIKFSYNADVIEREEMRLIQGHLKQALDCILTNPDVAVSDINIVPP
EEQVIOQLFNETERYVNTKIPOLFEQAKHTPEAAALMGNECYTQLOVRANOIA
HALIEKGVGSDIVAVMGRSMEMPAALLGIWKAGGAYMPLDPHPAERLSEFLKDSO
AAQLLIEDLISLPPSYEGNTITIEHTESYQTEAPNMPGDLAYLIYSGTGRPKG
VLVDHGLTANTLQMKREYSMTQDISLHLSFSYVDFGCVTSFLTLLSGACVLLTDD
EAKDVLAKRKRIARYKVSMMITVPSLYVLLVETADDAKSLRIVTFAGEAVTPDLLE
LNQIIPCASAEIANEYENSVATTILRLNKKERITIGHPTRINTKVFVLGNQMOP
GAAGELCISGAGLARGYKQOELTQAFSDHPFLERGLYRTGDAGRFLPDGTIEYIG
RFDQVKTGRYRIELREIETVLRQAPGYKEAVALRDYSAEKELVAVIPEKGNLP
DLIQLHATLSYMIPIASIIINISQPLISSGKLDRELPENNTSVIYMAPRTLIEA
DLAHIEDVNLQKHIGIKDDFFQGLGQSKAAALVSRHKKLNVELPLSEVFSYPTVE
SMAYKLSMKHAFTQIEPADQRDVYPLFSOKRLYALHQLADDSGYNMPAVLELRG
NLNORLSVLTENRHEALRTFVLDREDPVQIYYPMAFDFLKELEMESEQMLESA
IETFTKPYLSGGLPFRACVITMGNRGLFLLDHMHIIADGVSMSTLYOEFTDLYCGK
ELPALNHYKDFVWQOEKHPKELYKQOAYWLGOLGSLPTLEPLDKTRPLPDFR
GGTIEVNDKMADELHRLMAETGTLIYMLLAVYSIILLKSGQEDIVGSPAARPP
HADDERVIGMVTNLAMRSKPEGHFTSSYLHDIRHLALTAYEHQDYPFEELADKLDIT
NREVRNPLFAMLVLOSSEDFRFEVPGLSISSVTPKHDISKDFDLTHAEHLGIRC
RFXSTALFEETITQWASYFTELKGYTADTEMIRSNMQLLPAAERLLLEKMGQYA
AYPRNYSVFKEQVAYQPEHIAVCGHSOLTLYRDLNKAERAAAMLIKOCVRTGDI
VGLMDRSPMIGLISILKAGGAYLPIDPEYPERISFMLNDSGAKLLTERGLNKP
ADYTGHIYIDECENSIADVNIEIETDQAPVYIYTGTPGPKGVIEHRNVISL
LKQNLPEPHEDVWTLFHSYCFDSFVWEMFGALLNGSTLVVSKYARDQAFRL
LKKEVTVNPTAFYKLMLEQNDHNLNIRYVIFGGEALQPLQGLQSWNEKYPHTD
LINMYGITTETVHTFFKLAAADIAKNSNIGRPLSTLOAHVMDAHMNOPTGVPEGL
YIGGEGVARGYLNRDELADRFVSNPYLPDRLYRTGLAKRLSNGELEYLGRIDEQV
KVRHRIELGETOALLOYPMIKEAAVITRADEOQGTAYIYMWIKDOOAAANSIDRT
YLNKALPDFMLPAMQIDSIPTVTNGKLDOKALPEPEKQAYTADDISPRNEIETVMA
ETWELLNDELGVSAFNEFKLGGDSIKALQVCAKLRKQGFETTVREMEHQTLGELSA
RYKDVRAIDQGPVEGEITWPIQWFFSQSLESFFFNQSVMIYRAERFDEAALRKVL
KSLVTHHDALRIVRHEDGRQVINRGIDLDEELYALELDFDKDSLFEARNYIEEAA
SRMQCHILETGPLLHAGLFTENGDLHLFTIHLVVDVAVSWRILFEDFSTAYKQAVS
GESIKLPKQTSYLYTSQRIADYSISROVQREAAVWDECNHIOPIPKNDAAANTF
KDEVIDFELSRHHTPELLTAHKAYSTEMNDILLTALGLALQKWTGNQPKISMEGH
GRESVLEDIDISRTVGVWTSIYPPWLMRDSHDKKEERLGLIKQTKDMLHRIHPKG
AGYGVLYKISRWGSKNSPSELSFNGLQFODIQSNAPFVSDIKPGNEISPNWERPY
ALDTSAGVSSGCLNNHHIYNRFQFEKTIQTFSRHFQKTLENIIIEHCTKENQWWSAS
DFTDEDLDELSELTMGAVNKL"
17416...17421
17430...25097
/gene="pps3"
17430...25097
/gene="pps3"
/codon_start=1
/transl_table=11
/label=CDS5
/product="peptide-synthetase ORF3"
/protein_id="CAA84362.1"
/db_xref="GI:1805668"
/db_xref="SWISS-PROT:P39847"
/translation="MPQOPEIQDIYPLSFMOBGMFLHSLYDEQSRAYPEQASFTIHGO
LDLERFQKSDAVDFRDYDIFRTAFIYKNVAKPRQVVLKQCHPIHIEDLSHLNERDKE
HCTEAFKQDQSGDFDLQTDVLMRISILKAWPADHYVCIWSHHIIMDGWCUGIYIKDF
LHIYQALGKGQLPDLPPVQPYGTYIKWLMOQDREAAEYWKRLQHLQEKSTPLPKRTD

Query Match 5.9%; Score 59; DB 1; Length 39822;
Best Local Similarity 60.1%; Pred. No. 2.5e-06;
Matches 98; Conservative 0; Mismatches 65; Indels 0; Gaps 0;
QY 834 GTTGACATTACCGGATCCTCTTCAATCTCGGTGGACACTCTCTTGGCCACGAGCTC 893
DB 834 GTTGACATTACCGGATCCTCTTCAATCTCGGTGGGACACTCTCTTGGCCACGAGCTC 893
QY 894 ATTTCTCGTATCGACCAACGACTCAAGGTCCTGATCTGTCGTCGACGATCTCTTTGACCAT 953
DB 894 ATTTCTCGTATCGACCAACGACTCAAGGTCCTGATCTGTCGTCGACGATCTCTTTGACCAT 953
QY 954 CCTGTATTGCGGATCTAGCATCTGTCATCCGTCGACGGGCTG 996
DB 954 CCTGTATTGCGGATCTAGCATCTGTCATCCGTCGACGGGCTG 996
Search completed: June 4, 2003, 20:51:19
Job time: 2767 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2003, 17:28:42 ; Search time 271.5 Seconds
(without alignments)
8302.948 Million cell updates/sec

Title: US-09-482-788-1_COPY_7000_8000
Perfect score: 1001
Sequence: 1 aacctgtttcttaccctcg.....tcgtcaaggctgggtttg 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1010022.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1001	100.0	11212	21 AAA58762	DNA encoding a cyc
2	282.6	28.2	9633	22 AAF79702	Mycelia sterilia c
3	240.2	24.0	46899	15 AAQ54386	T. niveum Cyclospo
4	187.6	18.7	1713	15 AAQ54389	T. vasinfecta cycl
5	53	5.3	13029	23 AAS51470	Pseudomonas aerugi
6	50	5.0	41599	22 AAI66165	Bacillus subtilis
7	49.6	5.0	1200	20 AAV69559	Soil derived pepti
8	49.2	4.9	7178	21 AAA59145	DNA encoding a pep
9	48.2	4.8	37856	21 AAA11992	S. cellulosum DNA

10	48	4.8	18660	21 AAA58472	Nucleotide sequenc
11	46.8	4.7	6465	24 ABK74875	Bacillus lichenifo
12	46.8	4.7	8268	24 ABK74878	Bacillus lichenifo
13	45.4	4.5	4615	22 AAF90034	Nucleotide sequenc
14	45.4	4.5	34071	22 AAF90033	Nucleotide sequenc
c 15	45.4	4.5	42717	22 AAF90032	Nucleotide sequenc
16	42.8	4.3	88421	24 AAL40781	88421nt genomic DN
c 17	42	4.2	2365589	24 ABA90521	Genomic sequence o
18	41.8	4.2	58857	21 AAA58471	Nucleotide sequenc
19	40.4	4.0	7347	23 AAS54136	Pseudomonas aerugi
20	39.8	4.0	3798	24 ABK74884	Bacillus lichenifo
21	39.8	4.0	3798	24 ABK74902	Bacillus lichenifo
22	39.8	4.0	5451	22 AAF26319	Pseudomonas sp lip
23	39.4	3.9	68750	21 AAZ55887	Sorangium cellulos
24	39.4	3.9	71989	21 AAA29349	Sorangium cellulos
c 25	39.2	3.9	1172	20 AAV69527	Soil bacteria pept
c 26	39.2	3.9	4620	22 AAF26318	Pseudomonas sp lip
c 27	38.8	3.9	1178	20 AAV69562	Soil derived pepti
28	38.4	3.8	31122	14 AAQ40706	Bacillus subtilis
29	37.8	3.8	3849	22 AAF25795	S. chrysomallus ac
c 30	37.6	3.8	4403765	22 AAI99683	Mycobacterium tube
c 31	37.6	3.8	4411529	22 AAI99682	Mycobacterium tube
32	37.4	3.7	759	22 AAH52819	S. epidermidis ope
33	37.4	3.7	2301	22 AAH52513	S. epidermidis ope
c 34	37.4	3.7	2634	22 AAH54825	S. epidermidis gen
c 35	37.4	3.7	2744	22 AAH54705	S. epidermidis gen
c 36	37.4	3.7	3321	22 AAH54694	S. epidermidis gen
c 37	37.4	3.7	3602	22 AAH54066	S. epidermidis gen
c 38	37.4	3.7	3825	22 AAH54435	S. epidermidis gen
39	37.4	3.7	6968	22 AAH52883	S. epidermidis ope
40	36.8	3.7	1716	22 AAF26315	S. epidermidis ope
41	36.8	3.7	7110	22 AAF81361	Quorum sensing con
42	36.4	3.6	11601	12 AAQ13608	ACV synthetase gen
43	36.4	3.6	13058	14 AAQ48231	Vector containing
c 44	36.2	3.6	4541	20 AAV72100	Human RON receptor
c 45	36.2	3.6	4541	24 ABN97290	Gene #3788 used to

ALIGNMENTS

RESULT 1
AAA58762
ID AAA58762 standard; DNA; 11212 BP.
XX
AC AAA58762;
XX
DT 20-OCT-2000 (first entry)
XX
DE DNA encoding a cyclohexadepsipeptide synthetase.
XX
KW Cyclohexadepsipeptide synthetase; filamentous fungal cell;
KW cyclohexadepsipeptide; antibiotic; ss.
XX
OS Fusarium venenatum.
XX
FH Key Location/Qualifiers
FT CDS 270..9659
FT /*tag= a
FT /product= "cyclohexadepsipeptide synthetase"
XX
PN WO200042203-A2.
XX
PD 20-JUL-2000.
XX
PF 13-JAN-2000; 2000WO-US00913.
XX
PR 13-JAN-1999; 99US-0229862.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
XX
PI Berka RM, Rey MW, Yoder WT;
XX

DR WPI: 2000-482833/42.
DX P-PSDB: AAB07427.
XX
PT Producing a heterologous polypeptide for production of antibiotics
PT comprises cultivating a mutant of a parent filamentous fungal cell
PT comprising a nucleic acid sequence encoding cyclohexadepsipeptide -
XX
PS Claim 55; Fig 1A-I; 76pp; English.
XX
CC The present sequence encodes a cyclohexadepsipeptide synthetase
CC polypeptide. The specification describes a method for producing a
CC heterologous polypeptide. The method comprises cultivating a
CC mutant of a parent filamentous fungal cell, which produces less
CC cyclohexadepsipeptide than the parent filamentous fungal cell when
CC cultured under the same conditions. The method if used for the
CC production of biologically active compounds e.g. antibiotics.
XX
SQ Sequence 11212 BP; 2808 A; 2844 C; 2658 G; 2901 T; 1 other;
Query Match 100.0%; Score 1001; DB 21; Length 11212;
Best Local Similarity 100.0%; Pred. No. 4.9e-313;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AACCTGCTTTTCCACCTCGTTGAAGACAGAGTTTCCAGGCTGCTGGTGGACATCTTGAGA 60
Db 7000 AACCTGCTTTTCCACCTCGTTGAAGACAGAGTTTCCAGGCTGCTGGTGGACATCTTGAGA 7059
Qy 61 TCCTGCCAAAGAACATGGAAGCTGTGAATGAGCTCAGTGGGTATCGATATGCCGCTGTTG 120
Db 7060 TCCTGCCAAAGAACATGGAAGCTGTGAATGAGCTCAGTGGGTATCGATATGCCGCTGTTG 7119
Qy 121 TGCAGTTCGGGGTTCACCTGAGATGAGCTGTGTGCTCCGGTGGAAAGATGACTGGA 180
Db 7120 TGCAGTTCGGGGTTCACCTGAGATGAGCTGTGTGCTCCGGTGGAAAGATGACTGGA 7179
Qy 181 TCGACTTTCAAGGGAATCAATTGAACAGAGTCACTGGGTGACCTTCTCAAGTCTTCAG 240
Db 7180 TCGACTTTCAAGGGAATCAATTGAACAGAGTCACTGGGTGACCTTCTCAAGTCTTCAG 7239
Qy 241 ATGCTGCTATCGGAGTCAAGCAAAATTCCTTTCGAATACAGGCTTTCGAAGACAGG 300
Db 7240 ATGCTGCTATCGGAGTCAAGCAAAATTCCTTTCGAATACAGGCTTTCGAAGACAGG 7299
Qy 301 TCGTTCGCTTCCCTCAATAGCAACATCGATGAGTGGAGCTATCAACATTCGGTCCAGG 360
Db 7300 TCGTTCGCTTCCCTCAATAGCAACATCGATGAGTGGAGCTATCAACATTCGGTCCAGG 7359
Qy 361 CCGAGGCGGACTCATCACTATCCGTTCCCGACATCTTTCGATTCGCTGGGGAAGCCGGT 420
Db 7360 CCGAGGCGGACTCATCACTATCCGTTCCCGACATCTTTCGATTCGCTGGGGAAGCCGGT 7419
Qy 421 TCGGTGTCGAGGTCAGTTTCGACGACAGTGGTCTCAGATGATGGCATTCGACGCTGTTT 480
Db 7420 TCGGTGTCGAGGTCAGTTTCGACGACAGTGGTCTCAGATGATGGCATTCGACGCTGTTT 7479
Qy 481 TCCATCATCTTCTGCTCCAGGCGGTACTCTGCTCACTTTCTTACGGACCATCACCTTC 540
Db 7480 TCCATCATCTTCTGCTCCAGGCGGTACTCTGCTCACTTTCTTACGGACCATCACCTTC 7539
Qy 541 GAGGGTCTGATCTCTCCACCAATCGACCCCTTCAGGAGCTGCAAAACCGTCGATCGCCA 600
Db 7540 GAGGGTCTGATCTCTCCACCAATCGACCCCTTCAGGAGCTGCAAAACCGTCGATCGCCA 7599
Qy 601 TCGAAGTCCCGAGAGGCTTCGGTTCCTTACTTCCATCGTACATGATCCCATCGAATCG 660
Db 7600 TCGAAGTCCCGAGAGGCTTCGGTTCCTTACTTCCATCGTACATGATCCCATCGAATCG 7659
Qy 661 TTGTTCTGGACAGATGCCCTCTCAAGCCCAATGTAAGTTGACCGGAAGACTCTCTC 720
Db 7660 TTGTTCTGGACAGATGCCCTCTCAAGCCCAATGTAAGTTGACCGGAAGACTCTCTC 7719
Qy 721 GCAGGGCAAAAGTTGTACCGAAGCAGACAGCAGCGCGCTTACCGACATTTCCCATCA 780
Db 721 GCAGGGCAAAAGTTGTACCGAAGCAGACAGCAGCGCGCTTACCGACATTTCCCATCA 780

7720 GCAGGGCAAAAGTTGTACCGAAGCAGACAGCAGCGCGCTTACCGACATTTCCCATCA 7779
781 GTGAGGTGCAAGTCAATCTTTTCCGAAGAAGCACTGAGGTGTTGGCATGAAGTTGACA 840
7780 GTGAGGTGCAAGTCAATCTTTTCCGAAGAAGCACTGAGGTGTTGGCATGAAGTTGACA 7839
841 TTACCGATCACTTCTTCAATCTCGGTGGACACTCTCTCTTGGCCACGAAGCTCATTTCTC 900
7840 TTACCGATCACTTCTTCAATCTCGGTGGACACTCTCTCTTGGCCACGAAGCTCATTTCTC 7899
901 GTATCGCCACACGACTCAAGTCCGTATCACTGTCAAGGATGCTTTTGACCATCTGTAT 960
7900 GTATCGCCACACGACTCAAGTCCGTATCACTGTCAAGGATGCTTTTGACCATCTGTAT 7959
961 TTGCGGATCAGCATCTGTATCCGTCAAGGCTGGGTTG 1001
7960 TTGCGGATCAGCATCTGTATCCGTCAAGGCTGGGTTG 8000

RESULT 2
AAF79702
ID AAF79702 standard; DNA; 9633 BP.
XX AC AAF79702;
XX DT 29-MAY-2001 (first entry)
XX DE Mycellia sterilia cyclic depsipeptide synthase DNA.
XX KW Mycellia sterilia; cyclic depsipeptide synthase; anthelmintic;
XX KW PF1022; ds.
XX OS Mycellia sterilia.
XX PN WO200118179-A1.
XX PD 15-MAR-2001.
XX PF 07-SEP-2000; 2000WO-JP06103.
XX PR 07-SEP-1999; 99JP-0253040.
XX PR 06-APR-2000; 2000JP-0104291.
XX PA (MEIJ) MEIJI SEIKA KAISHA LTD.
XX PI Midoh N, Okakura K, Miyamoto K, Watanabe M, Yanai K, Yasutake T;
PI Aihara S, Futamura T, Kleinkauf H, Murakami T;
XX WPI: 2001-265970/27.
XX P-PSDB: AAB73958.
XX Novel cyclic depsipeptide synthase and gene encoding it for efficient
XX production of anthelmintic substance PF1022
XX Claim 3; Page 30-58; 92pp; Japanese.
XX The present sequence encodes a cyclic depsipeptide synthase from Mycellia
XX sterilia. The protein encoded by this sequence, or a protein containing
XX an addition, deletion and/or substitution of one or more amino acid
XX residues is useful for the efficient production of the anthelmintic
XX PF1022 (cyclo(D-lactyl-L-N-methylleucyl-D-3-phenyllactyl-L-N-
XX methylleucyl-D-lactyl-L-N-methylleucyl-D-3-phenyllactyl-L-N-
XX methylleucyl)).
XX SQ Sequence 9633 BP; 2318 A; 2834 C; 2462 G; 2019 T; 0 other;
Query Match 28.2%; Score 282.6; DB 22; Length 9633;
Best Local Similarity 59.5%; Pred. No. 3.2e-80;
Matches 618; Conservative 0; Mismatches 364; Indels 57; Gaps 6;
Qy 20 GTTCAAGACAGGTTTCCAGGCTGTTGGAAACATGTTGAGATCTGCTCCCAAGAACATGGA 79
Db 6900 GCTGCAGGCGCAGCTTCCCGATCGATCAAGCAGCTGAGATCCTCCGGAACATGCG 6959

Qy	121	TCGACGTTCCGGGTTTCAC	TGTGGAGATGAGCTGTG	TCTCCGGTTGAGAAAGATGACTGGA	180
Db	21196	TAGATCTGGCTGCTCGAGAGACAATCAACTATACACCAAGTCTCTCCCAACGCCTGGA	21255		
Qy	181	TCGACTTTCAAGCGAATCAAT	TGAACGAGAAGTCACTGGTGACCTTCTCAAGTCTTCAG	240	
Db	21256	TAGACTTTGGGCGAGACGGTCTCGACCGGAGACCCCTCATCAACTTGTGTGAAGGAGCACA	21315		
Qy	241	ATGCTGCTATCATGCGCAGTCA	CAGCAAAATTCCTTTGGAATCAGCGCTTTTAAACACAGG	300	
Db	21316	AGGATCCGGGACGCTCGCTATCGGTAATATCCCTGACAGCAAGACCATTTGTTGACGGT	21375		
Qy	301	TCGTCGTTCCCTCAATATGCAATCGACATCGATGAGTGGA	337		
Db	21376	TTGTCAACAAGTCACTGAGCGAGGATGATATGAGGAAGCGCAGAACTCACTGACCGAT	21435		
Qy	338	-----GCTATCAACCATTCGGTCCAGCGCGAGGGGACATCACTACTCCGTTCCCG	390		
Db	21436	CAGCTTGGGTTGAGCGCTCCGGATGGCGCTCAAAAGTGGCCCATCACTCGATGCAATGG	21495		
Qy	391	ACATCTTTCGCATTTGCTGGGGAAGCGGGTCCGTTCTGAGTCAAGTTCTGCACGACAGT	450		
Db	21496	ATGTCAAGGAGATTGCTCAGGAGCGGGATACCAAGTCAAGTCAAGTGGCGCGTCAAT	21555		
Qy	451	GGTCTCAGAATGGTGATTTGACGTGTTTTCATCATTTGTTGCTCCAAAGG	503		
Db	21556	GGTCCACAGAATGGTGGCTCGATGGCATCTTCCATCACTTCGAACCGCCCAAGGAGGTG	21615		
Qy	504	--CGTACTCTGGTCAACTTTCCTACGGACCATCACCTTCGAGGGTCTGATCTCTCCACCA	561		
Db	21616	CTGCGACACTTATTGAGTTTCCCGACGGATTAGCAAGCGCGGAATGTGAACACCTTTAACA	21675		
Qy	562	ATCGACCCCTTCAGGCACTTGCAAAACCGTCTGATCGCCATCGAAGTCCGCGCAGAGGCTTC	621		
Db	21676	ACGGTCCCTGACAGCATTTCAAGCGCCGCTCTTGGGACGCGAGATCCGCGAGAGTGC	21735		
Qy	622	GGTCTTACTTCCATCGTACATGATCCCATCGAATCGTTGTTCTGGACAAGATGCCTC	681		
Db	21736	AGACCCCTCTGCGGCTTACATGATCCCATCGCGCATCATGGTCTTGATCAGATGCGTG	21795		
Qy	682	TCAACGCCAATGGTAAAGTTGACCGGAGGAACTCTCTCGCAGGGCAAGGTTGTACCGA	741		
Db	21796	TCAACAAACAACGCGAAGATTGACCCGCAAGGAGCTTGTGCGGAGAGCTATCTGTGCCCCGA	21855		
Qy	742	AGCAGCAGACAGCAGCGCGTTTACCGACATTTCCCATCAGTGAAGTCAATCTTTT	801		
Db	21856	AGCCAAGGTCAGCGGCTAC---TCGGGTAGCCCCCGCAATGAGATCGAGGCTATTCTGA	21912		
Qy	802	GGGAGAAGCCACTGAGGTGTTGGCATGAAGTTGACATTCAGGATCACTTCTTCAATC	861		
Db	21913	GAGACGAATTCGAGGACGTGTCGGAACAGAGAATCAAGCTGCTGGATAACTTCTTTGATC	21972		
Qy	862	TCGGTGGACACTCTCTTTGGCCAGAGGATCAATTTCTCGTATCGACCAACGACTCAAGG	921		
Db	21973	TCGGGGGCACTCACTTATGGCCAGAGGCTCGCGCCCGCGTTAGCGCGCGCTTGATG	22032		
Qy	922	TCGGTACTCACTGAAGATGCTTTTGACCATCTCTGTATTTTTCGGGATCTAGCATCTGTCA	981		
Db	22033	CCCATATTTCCATCAAGATGCTTTGATCAGCCGGTCTGCGCGGATCTTTCGCGCTCCA	22092		
Qy	982	TCC 984			
Db	22093	TCC 22095			

RESULT 4
AAQ54389
ID AAQ54389 standard; DNA; 1713 BP.
XX
XX AAQ54389;
XX AC
XX AC
DT 08-JUL-1994 (first entry)
XX

DE	T. vasinfecta cyclosporin synthetase gene fragment.
XX	
KW	Enzyme; cyclosporin; synthetase-like activity; Tolypocladium niveum;
KW	T. inflatum GAMS; biosynthesis; vector; cyclosporin synthetase; ss.
XX	
OS	Tolypocladium vasinfecta.
XX	
PN	EP578616-A.
XX	
XX	12-JAN-1994.
XX	
PF	05-JUL-1993; 93EP-0810474.
XX	
XX	09-JUL-1992; 92AT-0001403.
PR	08-MAR-1993; 93AT-0000437.
PR	29-APR-1993; 93CH-0001310.
PR	04-MAY-1993; 93CH-0001375.
XX	
XX	(SANO) SANDOZ LTD.
PA	(SANO) SANDOZ PATENT GMBH.
PA	(SANO) SANDOZ-ERFINDUNGEN VERW GES MBH.
FA	
XX	
PI	Leitner E, Schneider E, Schoergendorfer K, Weber G;
XX	
DR	WPI; 1994-010432/02.
XX	
PT	Isolated DNA sequence - which codes for enzyme having cyclosporin
PT	synthetase like activity
XX	
PS	Example 16; Page 85; 93pp; English.
XX	
CC	This sequence represents a fragment of the cyclosporin synthetase
CC	gene from T. vasinfecta. This fragment was isolated using primers
CC	derived from the Tolypocladium niveum (formerly known as T. inflatum
CC	GAMS) cyclosporin synthetase gene, corresponding to bases 40309-40328
CC	and 42018-41999. The T. niveum gene encodes an enzyme which catalyses
CC	the peptide biosynthesis of cyclosporins and structurally related
CC	molecules. The T. niveum sequence may be used for the production of
CC	cyclosporin by transforming a vector containing the gene sequence in
CC	to a recombinant host. This allows effective production of antibiotic
CC	cyclosporin or its derivatives.
XX	
XX	Sequence 1713 BP; 390 A; 490 C; 469 G; 364 T; 0 other;

[illegible]

Db 1352 CCAAGCGCAAGGTGGATCGAAGAGTCGCTCGACAAAGCCCGGTCTATCCCAACAAATTG 1411
 QY 747 CAGACAGCAGCGCGCTTACCAGACATTTCCCATCAGTGAAGTCTGAAATC-ATTCTTTGCGA 805
 Db 1412 CCGAAGCAGCTTGGACTTTGTGGCGCCAGCAGCGAAATCGAGTCTGCTCTGCGA 1471
 QY 806 AGAGCCACTGAGGTGTTGGCATGAAGTTGACATTTACCGATCACTTCTTCAATCTCGG 865
 Db 1472 AGAATTTACCGATCTACTAGGCGTCAAGGTGGCATTTACAGACAACTTCTTCGAGTTGGG 1531
 QY 866 TGGACACTCTCTTGGCCAGCAAGCTCATTTCTCGTATCGACCAACCACTCAAGTCCG 925
 Db 1532 CGGCAATTCGCTTGGCCAGCAAGCTCAAGTCTGAGTCTTAAGTCCGACACTGGAGCGCGG 1591
 QY 936 TATCACTGTCAAGGATGCTTTTACCATCTCTGATTTTGGCATCTAGCATCTGTCATCCG 985
 Db 1592 TGTCACTGTGAAGCAGATCTTTGACCAAGCAGTACTTCTGCTGATCTTGTCTTATTCG 1651
 QY 986 TCAAGG 991
 Db 1652 TCAAGG 1657

RESULT 5

ID AAS51470 standard; DNA; 13029 BP.
 AC AAS51470;
 XX

DT 13-FEB-2002 (first entry)

DE Pseudomonas aeruginosa DNA for cellular proliferation protein #55.

KW Antisense; ds; prokaryotic cellular proliferation gene;
 KW antibiotic; antibacterial; drug design.
 OS

OS Pseudomonas aeruginosa.

PN WO200170955-A2.

PD 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-20727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

XX P-PSDB; ANU33611.

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

XX P-PSDB; ANU33611.

XX Claim 27; Seq ID NO 4052; 51pp; English.

XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the

CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence encodes an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX

SQ Sequence 13029 BP; 1978 A; 4677 C; 4485 G; 1889 T; 0 other;

Query Match 5.3%; Score 53; DB 23; Length 13029;

Best Local Similarity 51.9%; Pred No. 9.8e-06;

Matches 150; Conservative 0; Mismatches 130; Indels 9; Gaps 1;

QY 613 AGAGGCTTCGGTCTTACTTCCATCGTATCATCGATCCATCGAATCTGTTCTTGGACA 672

Db 12587 AGCAACTGCGCGCCGACCTGCGGACTACATGGTGGCTGCACTGGCTGCTGGACC 12646

QY 673 AGATGCTCTCAACGCCCAATGTAAGTTGACCGGAAGAACTCTTCGACGGCAAGG 732

Db 12647 GGATGCCCTCAACGCCCAAGCTCGACCGCAAGGCGCTGCCGCGTGGACATCG 12706

QY 733 TTGTACCGAAGCAGCAGACAGCAGCGCGTTTACCGACATTTCCCATCAGTGAAGTCTGAAG 792

Db 12707 GCCAGATCGAAGAACCCAGGCC-----TACCAGGCCCGCGCAACGACTGGAGGAAA 12757

QY 793 TCATTCTTTTGGGAAGACCACTGAGTGTGTTGGCATGAAGTGTGACATTTACCGATCACT 852

Db 12758 CCTTGGCGGATCTGCGCGAGGTGCTGAAGTCTGAGGCGGTGTTCTGACAACT 12817

QY 853 TCTTCAATCTCGTGGACACTCTCTCTTGGCCACGAAGTCTATTTCTCG 901

Db 12818 TCTTCAACTCGCGGGGCAATTCGTGCTGGCCACCCAGATCGCTCGCG 12866

RESULT 6

AAI66165

ID AAI66165 standard; DNA; 41599 BP.

AC AAI66165;

XX 15-JAN-2002 (first entry)

XX Bacillus subtilis Iturin A.

XX Bacillus subtilis; Iturin A; cation channel; ds.

XX Bacillus subtilis.

XX JP2001231561-A.

XX 28-AUG-2001.

XX 18-FEB-2000; 2000JP-0040825.

XX 18-FEB-2000; 2000JP-0040825.

XX (MASA/) MASADA M.

XX WPI; 2001-642167/74.

XX A gene encoding Iturin A for the production of large amounts of Iturin

PT A -

XX Disclos re; Page 4-17; 18pp; Japanese.

XX The invention relates to a gene encoding Iturin A encoding a protein

CC having an activity of promoting the transfer of a cation to the exterior

CC synthesis, modification and regulation of antibiotic Tel-aviv.
 CC Antibiotic Tel-aviv is useful in a wide range of clinical applications
 CC such as treating gingivitis. Antibiotic Tel-aviv is also useful for
 CC generating new biological agents from its secondary metabolites. The
 CC present sequence encodes a protein involved in synthesis of antibiotic
 CC Tel-aviv.

XX
 SQ Sequence 7178 BP; 1119 A; 2280 C; 2548 G; 1231 T; 0 other;

Query Match 4.9%; Score 49.2; DB 21; Length 7178;
 Best Local Similarity 50.9%; Pred. No. 0.00012;
 Matches 173; Conservative 0; Mismatches 158; Indels 9; Gaps 2;

QY 607 TCCGGAGAGGCTTCGGCTTACTTCCATCGTACATGCCATCGAATCGTGTGTC 666
 Db 2621 TCCGTGCACACCTCGCGGAAGTTCTCCCGACTACATGTCCTCCGCGACGCTTCGCGG 2680
 QY 667 TGGACAAGATGCTCTCAACGCCAATGTAAGTTGACCGGGAAGAACTCTTCGAGGG 726
 Db 2681 TGGATGCGATTCGCTGTGGGCAATGCAAGTGGACGGCGGCGAGCTGATGGCAGGC 2740
 QY 727 CAAGGTTGTACCGAAGCAGCAGCAGCGCCGTTACCGACATTTCCATCAGTGAGG 786
 Db 2741 C---GGTGTACCCGCGGGAAGACATCGCGGTCCATCCCGTCTGTGAGGCCA 2797
 QY 787 TCGAAGTCAATCTTTCGGAAGACCACTGAGGTGTTGGCATGAAGTTGACATTCACG 846
 Db 2798 CCTCGTCGAGCTGTGGAAGACGTCCTCCAGGT-----CAACGAGGTGGTGTGCGAGG 2851
 QY 847 ATCACTTCTCAATTCGGTGCACACTCTCTTCGCGCAGCAAGCTCATTTCTGATCG 906
 Db 2852 ATCGTCTTCTCAAGTGGGGGGGACTCGTCTGCGCCCGCTGGTGGAGGAGATGA 2911
 QY 907 ACCAAGCTCAAGTCCGTATCACTGTCAAGGATGCTT 946
 Db 2912 ACCGGGCTTCACAGCGGCTCGCGTCCACGACCTGTT 2951

RESULT 9

AA11992

ID AA11992 standard; DNA; 37856 BP.

XX AC AA11992;

XX DT 07-AUG-2000 (first entry)

XX S. cellulosum DNA encoding polyketide and heteropolyketide enzymes.

XX Polyketide biosynthesis; heteropolyketide biosynthesis; mutasynthesis;

KW epothilone; cytotoxic; immunosuppressant; antibiotic; antifungal;

KW plant-protection; ds.

XX OS Sorangium cellulosum.

XX FH Location/Qualifiers

FT CDS complement (3398..6100)

FT /tag= a

FT /product= "ORF1-trna synthetase"

FT /note= "gtg start codon"

FT complement (6374..7111)

FT /tag= b

FT /product= "ORF2-monoxygenase"

FT complement (8433..9550)

FT /tag= c

FT /product= "ORF3-aminotransferase"

FT /note= "AGT start codon given in the specification"

FT 9855..11393

FT /tag= d

FT /product= "ORF4- tyrosine/DOFA-Decarboxylase"

FT /note= "GTG start codon"

FT 12212..13658

FT /tag= e

FT /product= "ORF5-3-oxoacyl-ACP-reductase"

FT CDS /note= "ACC start codon"

FT 15374..19984

FT /tag= f

FT /product= "ORF6-polyketide synthase"

FT 20003..27889

FT /tag= g

FT /product= "ORF7-peptide synthetase"

FT 28251..29400

FT /tag= h

FT /product= "ORF8-transpeptidase"

FT complement (30040..31720)

FT /tag= i

FT /product= "ORF9-regulation element"

FT /note= "CGC stop codon"

FT 31982..32932

FT /tag= j

FT /product= "ORF10-transcription regulator"

FT 33128..33613

FT /tag= k

FT /product= "ORF11-regulation element"

FT /note= "GTG start codon"

FT 33661..34077

FT /tag= l

FT /product= "ORF12-regulation element"

FT complement (35255..35616)

FT /tag= m

FT /product= "ORF13-transcription regulator"

FT complement (35730..36242)

FT /tag= n

FT /product= "ORF14-transcription regulator"

FT /note= "GTG start codon"

XX DE19846493-Al.

XX 13-APR-2000.

XX 09-OCT-1998; 98DE-1046493.

XX 09-OCT-1998; 98DE-1046493.

XX (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.

XX Beyer S, Mueller R;

XX WPI; 2000-294101/26.

XX DNA sequence coding for products involved in the biosynthesis of

XX polyketide or heteropolyketide compounds, especially epothilone

XX Claim 3; Page 20-33; 36pp; German.

XX This invention describes a novel DNA sequence (I) whose expression

XX products effect or are involved in the enzymatic biosynthesis,

XX mutasynthesis or partial synthesis of polyketide or heteropolyketide

XX compounds (ii). (I) can be inserted into an expression vector and used

XX to transform or transfect prokaryotic or eukaryotic cells with the aim

XX of obtaining strains that produce large amounts of polyketide or

XX heteropolyketide compounds, especially epothilones, which have cytotoxic

XX and/or immunosuppressant and antibiotic and antifungal activities and

XX are useful as plant-protection agents. This sequence represents the DNA

XX sequence isolated from Sorangium cellulosum which is described in the

XX method of the invention.

XX Query Match 4.8%; Score 48.2; DB 21; Length 37856;

XX Best Local Similarity 58.9%; Pred. No. 0.00063;

XX Matches 83; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 830 GAAGGTTGACATTACCGATCATTCTCAATCTCGGTGGACACTCTCTTGGCCACGAA 889

Db 19714 GCAGGTAGGCGCAACGACGATTTCTCCAGTGGCGGCCATTCGCTGTTGCCACGCA 19773

XX AAF90034;
AC
XX
DT DT 06-AUG-2001 (first entry)
XX
DE Nucleotide sequence of a type I polyketide synthase.
XX
KW Metabolic pathway operon; polyketide; polyketide antibiotic;
KW type I polyketide synthase; ss.
OS Unidentified.

FH Key Location/Qualifiers
FT CDS 2..4615
FT /*tag= a
FT /product= "type I polyketide synthase"
XX

PN W0200140497-A2;
XX
PD 07-JUN-2001.
XX
PF 27-NOV-2000; 2000WO-FR03311.
XX
PR 29-NOV-1999; 99FR-0015032.
PR 07-JUN-2000; 2000US-0209800.
XX
PA (AVET) AVENTIS PHARMA SA.
XX
PI Jeannin P, Pernodet J, Guerinneau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX
DR WPI; 2001-374849/39.
DR P-PSDB; AAB83971.
XX

PT Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide syntheses and derived
PT antibiotics
XX
PS Claim 35; Page 300-302; 356pp; French.
XX

CC The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. AAF90034-39
CC represent open reading frames (ORFs) of the coding strand of cosmid
CC a26g1, and encode type I polyketide syntheses.
XX

SQ Sequence 4615 BP; 827 A; 1477 C; 1429 G; 882 T; 0 other;
Query Match 4.5%; Score 45.4; DB 22; Length 4615;
Best Local Similarity 49.8%; Pred. No. 0.0016;
Matches 146; Conservative 0; Mismatches 141; Indels 6; Gaps 1;

QY 630 CTTCATCGTACATCATCCATCGAACATCTGTTCTGGACAAGATGCCCTCAAGCC 689
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4157 CTGCCGACTACATGATCCGACCGTGCTGTCTGCACGAATAATGCCGCTGACGCC 4216
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 690 AATGTTAAAGTTGACCGGAAGAAGTAATCTCTCGCAGGGCAAAGTTGTACCGAAGCAGCAG 749
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4217 AACGGAATAATGCACCGTAACGCCCTGCGCATCCCGAGCCAGCCGCGGAGCCAGCC 4276
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 750 ACAGCAGCGCGTTACCAGACATTTCCCATCATGAGTGGAAGTCAATTCTTTCGAAGAA 809
|||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4277 GAAGCATTTACGCGCTCCGGAATACTCCGTTGGAACAGGTAATCGCCCCACATTTGGGGCGAG 4336
|||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: June 4, 2003, 19:04:47 ; Search time 57 Seconds
(without alignments)
5385.674 Million cell updates/sec

Title: US-09-482-788-1_COPY_7000_8000

Perfect score: 1001

Sequence: 1 aactgtttttttacacgcg.....tccgtcaaggctgggttg 1001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	240.2	24.0	46899	1 US-08-471-119A-1	Sequence 1, Appl
2	189.2	18.9	1713	1 US-08-471-119A-4	Sequence 4, Appl
3	49.6	5.0	1200	4 US-08-861-774E-85	Sequence 85, Appl
4	39.4	3.9	68750	3 US-09-335-409-1	Sequence 1, Appl
5	39.4	3.9	68750	4 US-09-568-102-1	Sequence 1, Appl
6	39.4	3.9	68750	4 US-09-567-969-1	Sequence 1, Appl
7	39.4	3.9	68750	4 US-09-568-480-1	Sequence 1, Appl
8	39.4	3.9	68750	4 US-09-568-486-1	Sequence 1, Appl
9	39.4	3.9	68750	4 US-09-568-472-1	Sequence 1, Appl
10	39.4	3.9	68750	4 US-09-567-899-1	Sequence 1, Appl
11	39.4	3.9	71989	4 US-09-443-501A-2	Sequence 2, Appl
12	39.2	3.9	1172	4 US-08-861-774E-17	Sequence 17, Appl
13	38.8	3.9	1178	4 US-08-861-774E-91	Sequence 91, Appl
14	38.8	3.9	7218	1 US-08-232-463-14	Sequence 14, Appl
15	37.6	3.8	4403765	4 US-09-103-840A-2	Sequence 2, Appl
16	37.6	3.8	4411529	4 US-09-103-840A-1	Sequence 2, Appl
17	36.4	3.6	11601	2 US-08-222-617A-3	Sequence 3, Appl
18	36.4	3.6	11601	2 US-08-222-617A-24	Sequence 24, Appl
19	35.8	3.6	7215	4 US-09-134-001C-627	Sequence 627, App
20	34.6	3.5	1177	4 US-08-861-774E-23	Sequence 23, Appl
21	33.4	3.3	4220	4 US-09-183-846A-11	Sequence 11, Appl
22	33.4	3.3	4220	4 US-08-961-578C-11	Sequence 11, Appl
23	33.2	3.3	527	1 US-07-998-973A-22	Sequence 22, Appl
24	33.2	3.3	527	1 US-08-452-800-22	Sequence 22, Appl
25	33.2	3.3	527	5 PCT-US92-11353-22	Sequence 22, Appl
26	33.2	3.3	2620	1 US-08-056-200-109	Sequence 109, App
27	33.2	3.3	2620	2 US-08-800-644-109	Sequence 109, App

Sequence 93, Appl
Sequence 167, App
Sequence 3, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 22, App
Sequence 83, Appl
Sequence 6, Appl
Sequence 81, Appl
Sequence 2065, Ap
Sequence 64, Appl
Sequence 22, App
Sequence 166, App
Sequence 26, Appl
Sequence 1, Appl
Sequence 72, Appl
Sequence 1, Appl
Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-471-119A-1
; Sequence 1, Application US/08471119A
; Patent No. 5827706
; GENERAL INFORMATION:
; APPLICANT: Leitner, Ernst
; APPLICANT: Schneider, Elisabeth
; APPLICANT: Schoergendorfer, Kurt
; APPLICANT: Weber, Gerhard
; TITLE OF INVENTION: Cyclosporin Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5827706artis Corporation
; STREET: 59 Route 10
; CITY: East Hanover
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07936
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,119A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kasenoff, Melvyn
; REGISTRATION NUMBER: 26,389
; REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 503 8474
; TELEFAX: 201 503 8807
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46899 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Tolypocladium niveum
; STRAIN: ATCC 34921
US-08-471-119A-1

Query Match 24.0%; Score 240.2; DB 1; Length 46899;
Best Local Similarity 55.0%; Pred. No. 1.1e-69;

```

RESULT 2
US-08-471-119A-4
: Sequence 4, Application US/08471119A
: Patent No. 5827706
: GENERAL INFORMATION:
: APPLICANT: Leitner, Ernst
: APPLICANT: Schneider, Elisabeth
: APPLICANT: Schoengendorfer, Kurt
: APPLICANT: Weber, Gerhard
: TITLE OF INVENTION: Cyclosporin Synthetase
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 5827706artis Corporation
: STREET: 59 Route 10
: CITY: East Hanover
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07936
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/471.119A
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Kassenoff, Melvyn
: REGISTRATION NUMBER: 26,389
: REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201 503 8474
: TELEFAX: 201 503 8807
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1713 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Necosmospora vasinfecta
: US-08-471-119A-4

Query Match 18.9%; Score 189.2; DB 1; Length 1713;
Best Local similarity 59.5%; Pred. No. 2.2e-53;
Matches 396; Conservative 0; Mismatches 258; Indels 12; Gaps 4;

QY 338 GCTATCAACCATTCGGTCAGCCGCCAGGCGGACTATCACTATCCGTTCCCGACATCTT 397
Db 992 GATATCGGCCACTCAATACACGAGGGAAGTAATGCGCTGCTCTCTCAGTGGCGGACCTGAT 1051
QY 398 TCGCATTTGCTGGGGAAGCCGGGTTCGGTGTGAGGTCAGTTCTGCAGCAGACAGTGGTCTCA 457
Db 1052 TGAGATTGTGAAGGGATCGGCTTCCAAAGTTGAGACCAAGCTGGGCTCGACAACACTCCA 1111
QY 458 GAATGGTGCATTGGACGCTGTTTTTCATCATATTTGTTGGTCCCAA-----GGGCGTAC 508
Db 1112 GCGCGCGGCACTCGATGCTGTTTTTCCACCGAATTCGAAAACCAAGACACTCGGGGTCA 1171
QY 509 TCTGGTCAACTTTCCTACGAGCAATCACTTCGAGGGTCTGTGATCTC-CTCACCACATCGAC 567
Db 1172 CATGTTTCAGTTCCTCCAACTGAACACAAGGGCGGTCTTCGACAGCAGTCTCACGAATCGCC 1231
QY 568 CCCTTCAGCGACTGCAAAACCGTGTGATCGCCATCGAAAGTCCGCGAGAGGCTTCGGTCTCT 627
Db 1232 CCCTACACCTGTTTCAGAGCCGCGGCTCGAGGCAAGGTCGCGAGCGGCTGCAATCGC 1291

```

```

QY 628 TACTTCCATCGTACATGATCCATCGAACATCGTTGTTCTGACAAAGATGCTCTCAAG 687
    |||||
Db 1292 TCGTTCATCGTACATGATTCCTCTCGGATCATGTTGCTCGATCAGATGCTCTCACGT 1351
    |||||
QY 688 CCAATGGTAAATGACCGAGGAACTCTCTCGCAGGCGCAAGGTTGTACCGA-AGCAG 746
    |||||
Db 1352 CCAACGGCAAGGTGGATCGCAAGAAGCTCGTCGACAGCCCGGGTCAATCCCAACAATG 1411
    |||||
QY 747 CAGACAGACGCGCGTACCGACATTTCCCATCAGTGAGGTGGAAGTC-ATTCTTTGCGA 805
    |||||
Db 1412 CGCGAAGCAGCTTGACATTTGTCGCGCGCACCGAATCGAGTCTGCTCGCA 1471
    |||||
QY 806 AGAAGCCATGAGGTGTTGGCATGAAGTTGACATTAACCGATCACTTCTTCAATCTGG 865
    |||||
Db 1472 AGAATTTACCGATCTACTAGGCGTCAAGGTGCGCATTAACAGCAACTTCTTCAGTTGG 1531
    |||||
QY 866 TGGACACTCTCTTGGCCACCAAGCTCATTTCTCGTATCGACCAAGACTCAAGTCCG 925
    |||||
Db 1532 CGGCCATTCGCTGCTGGCCACCAAGCTGAGCGCAGCTCTAAAGTCGACACTGGAGCGCG 1591
    |||||
QY 926 TATCACTGTCAAGGATGCTTTGACCATCTCTGATTTGCGGATCTAGCATCTGTCATCG 985
    |||||
Db 1592 TGTCACTGTGAAGCAGATCTTTGACCGCAGTCTTGTGATCTGCTGATCTGCTTCTATT 1651
    |||||
QY 986 TCAAGG 991
    |||||
Db 1652 TCAAGG 1657
    |||||

```

RESULT 3

```

US-08-861-774E-85
; Sequence 85, Application US/08861774E
; Patent No. 6297007
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/08/861,774E
; CURRENT FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 85
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Clone ps30
US-08-861-774E-85

```

```

Query Match 5.0%; Score 49.6; DB 4; Length 1200;
Best Local Similarity 63.3%; Pred. No. 1.8e-06;
Matches 76; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 596 CGCCATCGAAGTCCGCGAGAGGCTTCGGTCTTACTTCCATCGTACATGATCCCATCGAA 655
    |||||
Db 928 CGACTCGGAGCTGCGGAGTTCTCTGGGACGTACTCCCGGAGCGGATGATACCTCGGC 987
    |||||
QY 656 CATCGTGTTCGGACAAGATGCTCTCAACGCCAATGTAAGTTGACCGGAAGAACT 715
    |||||
Db 988 ATTCGTTGTGCTGGAGACGCTCCCACTGACCCACAACGGAAGGTGGACCGAGAGCCCT 1047
    |||||

```

RESULT 4

```

US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James

```

```

; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkie, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match 3.9%; Score 39.4; DB 3; Length 68750;
Best Local Similarity 57.9%; Pred. No. 0.059;
Matches 70; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 595 TCGCCATCGAAGTCCGCGAGAGGCTTCGGTCTTACTTCCATCGTACATGCCCATCGA 654
    |||||
Db 15638 TCGCGATATGCTTCGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTCCCTACAG 15697
    |||||
QY 655 ACATCGTTCCTTGGACAAGATGCCCTCTCAACGCCAATGTAAGTTGACCGGAAGAAC 714
    |||||
Db 15698 TCTTCGTGGAGTTCGATGCGTTGCCGCTGACGTCCAAACGCAAGTTCGATCGTAAGGCC 15757
    |||||
QY 715 T 715
    ||
Db 15758 T 15758
    ||

```

RESULT 5

```

US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkie, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1

```

```

Query Match 3.9%; Score 39.4; DB 4; Length 68750;
Best Local Similarity 57.9%; Pred. No. 0.059;
Matches 70; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 595 TCGCCATCGAAGTCCGCGAGAGGCTTCGGTCTTACTTCCATCGTACATGCCCATCGA 654
    |||||
Db 15638 TCGCGATATGCTTCGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTCCCTACAG 15697
    |||||
QY 655 ACATCGTTCCTTGGACAAGATGCCCTCTCAACGCCAATGTAAGTTGACCGGAAGAAC 714
    |||||
Db 15698 TCTTCGTGGAGTTCGATGCGTTGCCGCTGACGTCCAAACGCAAGTTCGATCGTAAGGCC 15757
    |||||
QY 715 T 715
    ||
Db 15758 T 15758
    ||

```

```
RESULT 6
US-09-567-969-1
: Sequence 1, Application US/09567969
: Patent No. 6355457
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/567,969
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-567-969-1

Query Match      3.9%; Score 39.4; DB 4; Length 68750;
Best Local Similarity 57.9%; Pred. No. 0.059;
Matches 70; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 595 TCGCCATCGAAGTCCGGCAGAGGCTTCGGTCTTACTTCCATCGTACATGATCCCATCGA 654
    |||||
Db 15638 TCGCCGATATGCTTCGGGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAG 15697

QY 655 ACATCGTTGTTCTGGACAAGATGCTCTCAACGCCAATGTTAAAGTTGACCGGAAGAAC 714
    |||||
Db 15698 TCITCGTGGAGCTCGATGCGTTCGCGCTGAGTCCAAACGGCAAGTCTGATCGTAAGGCC 15757

QY 715 T 715
Db 15758 T 15758

RESULT 7
US-09-568-480-1
: Sequence 1, Application US/09568480
: Patent No. 6355458
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,480
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-568-480-1

Query Match      3.9%; Score 39.4; DB 4; Length 68750;
Best Local Similarity 57.9%; Pred. No. 0.059;
Matches 70; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 595 TCGCCATCGAAGTCCGGCAGAGGCTTCGGTCTTACTTCCATCGTACATGATCCCATCGA 654
    |||||
Db 15638 TCGCCGATATGCTTCGGGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAG 15697

QY 655 ACATCGTTGTTCTGGACAAGATGCTCTCAACGCCAATGTTAAAGTTGACCGGAAGAAC 714
    |||||
Db 15698 TCITCGTGGAGCTCGATGCGTTCGCGCTGAGTCCAAACGGCAAGTCTGATCGTAAGGCC 15757

QY 715 T 715
Db 15758 T 15758

RESULT 8
US-09-568-486-1
: Sequence 1, Application US/09568486
: Patent No. 6355459
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,486
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match      3.9%; Score 39.4; DB 4; Length 68750;
Best Local Similarity 57.9%; Pred. No. 0.059;
Matches 70; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 595 TCGCCATCGAAGTCCGGCAGAGGCTTCGGTCTTACTTCCATCGTACATGATCCCATCGA 654
    |||||
Db 15638 TCGCCGATATGCTTCGGGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAG 15697

QY 655 ACATCGTTGTTCTGGACAAGATGCTCTCAACGCCAATGTTAAAGTTGACCGGAAGAAC 714
    |||||
Db 15698 TCITCGTGGAGCTCGATGCGTTCGCGCTGAGTCCAAACGGCAAGTCTGATCGTAAGGCC 15757

QY 715 T 715
Db 15758 T 15758

RESULT 9
US-09-568-472-1
: Sequence 1, Application US/09568472
: Patent No. 6358719
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,472
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
```


Db 273 CGCGTCGAGTGGCGGATCATCTTCGACGCGGTTCCCGAGTACATGATTCGGGTGC 214
Qy 656 CATCTGTTGTTCTGACAAAGATGCTCTCAACGCAATGGTAAAGTTGACCGGAAGAACT 715
Db 213 ATTCTGAGCTCCGCGAGATGCGCGCATGGCGGAGCGCAAGATCGATCTGAAGGTGCT 154
Qy 716 CTCT 719
Db 153 GCCT 150

RESULT 13

US-08-861-774E-91/c
Sequence 91, Application US/08861774E

Patent No. 6297007

GENERAL INFORMATION:

APPLICANT: Waters, Barbara

APPLICANT: Miao, Vivian

APPLICANT: Ho, Yip

APPLICANT: Tong, Seow

TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR

TITLE OF INVENTION: BIOACTIVE MOLECULES

FILE REFERENCE: 9993-006

CURRENT APPLICATION NUMBER: US/08/861,774E

CURRENT FILING DATE: 1997-05-22

NUMBER OF SEQ ID NOS: 94

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 91

LENGTH: 1178

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Clone ps25

US-08-861-774E-91

Query Match:

Best Local Similarity 58.8%; Score 38.8; DB 4; Length 1178;

Matches 67; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 602 CGAAGTCGCGAGAGGCTCGGTCCTTACTTCCATCGTACATGATCCCATCGAACATCGT 661

Db 270 CGAGTCCGCGAGCGGCTCGCGCGAGACTCCCGGAGTACATGTCGACAGGTGATCGT 211

Qy 662 TGTTCTGGACAAGATGCCTCTCAACGCCATGTAAAGTTGACCGGAAGAACT 715

Db 210 ATCGCTCGCGCGCTGCGCTCACGCGGACGACAAAGATCGACCGCAAGGCCCT 157

RESULT 14

US-08-232-463-14

Sequence 14, Application US/08232463

Patent No. 5670367

GENERAL INFORMATION:

APPLICANT: DORNER, F.

APPLICANT: SCHEIFLINGER, F.

APPLICANT: FALKNER, F. G.

TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 1800 Diagonal Road, Suite 500

CITY: Alexandria

STATE: VA

COUNTRY: USA

ZIP: 22313-0299

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/232,463

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION NUMBER:

FILING DATE: US/07/935,313

APPLICATION NUMBER: EP 91 114 300.6

FILING DATE: 26-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472/114 IMMU

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)836-9300

TELEFAX: (703)683-4109

TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 7218 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: pTZgpt-F1s

US-08-232-463-14

Query Match

Best Local Similarity 2.6%; Score 38.8; DB 1; Length 7218;

Matches 10; Conservative 210; Mismatches 162; Indels 0; Gaps 0;

Qy 301 TCGTCGCTCCCTCAATAGCAACATCGATGAGTGGCAGCTATCAACCATTCGGTCCAGCG 360

Db 1068 YY 1127

Qy 361 CCGAGGCGAGTCACTACATCGCTCCCGACATCTTCGCATTCCTGGGAAGCCGGT 420

Db 1128 YY 1187

Qy 421 TCGGTGCGAGTCACTTCGACGAGTGTCTCAGATGTTGACGCTGTTT 480

Db 1188 YY 1247

Qy 481 TCCATCATCTGTCCTCCCAAGGCGTACTCTGGTCAACTTTCCTACGACCATCACCTTC 540

Db 1248 YY 1307

Qy 541 GAGGCTGTGATCTCCTCAACCAATCGACCCCTTCAGCGACTGCAAAACCGTGTATCGCCA 600

Db 1308 YY 1367

Qy 601 TCGAAGTCCCGAGAGGCTTCGGTCTTACTTCATCGTACATGATCCCATCGAACATCG 660

Db 1368 YY 1427

Qy 661 TTGTTCTGGACAAGATGCCTCT 682

Db 1428 YYYYYYGTACCAAAATCTTCT 1449

RESULT 15

US-09-103-840A-2/c

Sequence 2, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      3.8%; Score 37.6; DB 4; Length 4403765;
Best Local Similarity 61.0%; Pred. No. 3.1;
Matches 61; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY      607 TCCGCGAGAGGCTTCGGTCTTACTTCCATCGTACATGATCCGATCGAAGATCGTTGTTTC 666
        ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      2661750 TCCGCGCCGGTGGCGCGGCGACTGCTGAATATATCTGCCCGCGCCTATGTGTGC 2661691

QY      667 TGGACAAGATGCCCTCTCAACGCCAATGGTAAAGTTGACCG 706
        ||| |||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      2661690 TGGATGAGATTCCGATCACCGCGCATGGCAAAATCGACCG 2661651
```

Search completed: June 4, 2003, 21:51:15
Job time : 74 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2003, 20:51:29 ; Search time 170 Seconds
(without alignments)
8240.805 Million cell updates/sec

Title: US-09-482-788-1_COPY_7000_8000
Perfect score: 1001
Sequence: 1 aactgtcttttccactcg.....tcgtcaaggctgggttg 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 870385 seqs, 699768693 residues

Total number of hits satisfying chosen parameters: 1740770

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	53	5.3	13029	10	US-09-815-242-4052
2	49.6	5.0	1200	10	US-09-924-256A-85
3	46.8	4.7	6465	10	US-09-974-300-2166
4	46.8	4.7	8268	10	US-09-974-300-2169
5	42.8	4.3	88421	9	US-09-976-059-1
6	40.4	4.0	7347	10	US-09-815-242-7773
7	39.8	4.0	3798	10	US-09-974-300-2175
8	39.8	4.0	3798	10	US-09-974-300-2193
9	39.4	3.9	68750	9	US-10-014-717-1
10	39.2	3.9	1172	10	US-09-924-256A-17
11	38.8	3.9	1178	10	US-09-924-256A-91
12	38	3.8	1160	9	US-10-184-634-592
13	37.2	3.7	499	9	US-10-184-644-592
14	37.2	3.7	499	9	US-10-184-634-592
15	36.4	3.6	513	9	US-10-123-155-536
16	36.2	3.6	547	9	US-10-066-543-1765
17	36.2	3.6	4541	9	US-10-123-036-3
18	35.2	3.6	4541	10	US-09-880-107-3785
19	35.6	3.6	290	10	US-09-294-093B-2985

20	35.6	3.6	1803	9	US-10-166-087-43	Sequence 43, Appl
21	35.6	3.6	32539	9	US-10-166-087-1	Sequence 1, Appl
22	35.4	3.5	3471	10	US-09-974-300-2167	Sequence 2167, Ap
23	35.2	3.5	270	7	US-10-060-036-2566	Sequence 2566, Ap
24	35.2	3.5	400	7	US-08-781-986A-3817	Sequence 3817, Ap
25	35.2	3.5	400	7	US-08-781-986A-3926	Sequence 3926, Ap
26	35.2	3.5	29555	7	US-08-781-986A-206	Sequence 206, App
27	34.6	3.5	716	9	US-10-123-155-96	Sequence 23, Appl
28	34.6	3.5	1177	10	US-09-924-256A-23	Sequence 346, App
29	34	3.4	671	9	US-10-184-644-346	Sequence 346, App
30	34	3.4	671	9	US-10-184-634-346	Sequence 134, App
31	33.6	3.4	518	10	US-09-833-381-134	Sequence 134, App
32	33.6	3.4	518	10	US-09-833-381-135	Sequence 135, App
33	33.4	3.3	244	10	US-09-923-876-191	Sequence 191, App
34	33.4	3.3	410	9	US-09-918-995-6617	Sequence 6617, Ap
35	33.4	3.3	4157	7	US-08-556-422-1	Sequence 1, Appli
36	33.4	3.3	7158	10	US-09-974-300-2171	Sequence 2171, Ap
37	33.2	3.3	445	9	US-09-918-995-21047	Sequence 21047, A
38	33.2	3.3	487	10	US-09-833-381-1125	Sequence 1125, Ap
39	33.2	3.3	487	10	US-09-833-381-1126	Sequence 1126, Ap
40	33.2	3.3	599	9	US-10-198-846-9039	Sequence 9039, Ap
41	33.2	3.3	2619	10	US-09-962-832-223	Sequence 223, App
42	32.8	3.3	1178	10	US-09-924-256A-93	Sequence 93, Appl
43	32.6	3.3	510	9	US-10-123-155-112	Sequence 112, App
44	32.6	3.3	1188	9	US-09-764-868-433	Sequence 433, App
45	32.6	3.3	1551	10	US-09-815-242-9546	Sequence 9546, Ap

ALIGNMENTS

RESULT 1
US-09-815-242-4052
Sequence 4052, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.01A
CURRENT FILING DATE: 2001-03-21
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4052
LENGTH: 13029
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4052

Query Match 5.3%; Score 53; DB 10; Length 13029;
Best Local Similarity 51.9%; Pred. No. 7.9e-07;

Matches	150;	Conservative	0;	Mismatches	130;	Indels	9;	Gaps	1;
QY	613	AGAGGCTCGGTCTTACTTCCATCGTACATGATCCATCGAACATCGTTGTCTGGACA							672
Db	12587	AGCAACTGGCGCGCACTTCCGCGGACTACATGGTCCGCTGCACCTGGCTGTGGACC							12646
QY	673	AGATGCTCTCAAGCGCAATGTAAAGTTGACCGGAAGGAACTCTTCGCAAGGCAAAG							732
Db	12647	GSATGCCGTCAACGCGCAACGCAAGCTCGACCGCAAGCGCTGCCGCGTGGACATCG							12706
QY	733	TTGTACCGAAGCAGCAGACAGCAGCGCGGTTTACCGACATTTCCCATGCTGAGTCTGAAG							792
Db	12707	GCCAGATGCAGAACCCAGGCC-----TACAGGCCCGCGCAACCACTGGAGGAAA							12757
QY	793	TCATTCTTTGCGAAGAAGCCACTGAGGTGCTTTGGCATGAAGGTTGACATFACCAGTCACT							852
Db	12758	CCCTGGCGGGATCTGGCCGAGGTCTGAAGGTTCAGCGGCTGGGGGTGTTCGACAACT							12817
QY	853	TCCTCAATCTCGGTGACACTCTCTCTTTGGCCACGAAGCTCATTTCTCG							901
Db	12818	TCCTCGAACTCGCGGGGATTCGCTGCTGGCCACCCAGATCGCCTCGCG							12866

```

RESULT 2
US-09-924-256A-85
; Sequence 85, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/09/924,256A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 08/861,774
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 85
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Clone ps30
US-09-924-256A-85

```

	Query Match	5.0%	Score 49.6;	DB 10;	Length 1200;		
	Best Local Similarity	63.3%;	Pred. No. 3e-06;				
Matches	76;	Conservative	0;	Mismatches	44; Indels	0; Gaps	0;
QY	596	CGCCATCGAAGTCCGGGAGAGGTTTCGGTCTTTACTTCCATCGTACATGATCCCATCGAA	655				
Dd	928	CGACTCGGAGCTGCGGAGTTCCTGCGGACGTTCCTCCCGAGCGGATGATACCTCGGC	987				
QY	656	CATCTGTTGTTCTGGACAAGATGCTCTCAACGCCATGTTAAAGTTGACC CGGAAGNACT	715				
Dd	988	ATTCTGTTGTCTGGAGACGCTCCACTGACCCACAACGGGAAGGTGACCGACAGAGCCCT	1047				

```

RESULT 3
US-09-974-300-2166
; Sequence 2166, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression.
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER:US/09/974,300

```

```

; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2166
; LENGTH: 6465
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2166

```

	Query Match	4.7%	Score 46.8;	DB 10;	Length 6465;
	Best Local Similarity	63.2%;	Pred. No. 7.4e-05;		
	Matches	72;	Conservative	0;	Mismatches 42; Indels 0; Gaps 0;
QY	602	CGAAGTCCGCGAGAGGTTCCGGTCTTACTTCATCGTACATGATGCCATCGCAACATCGT	661		
Db	3662	CGAAATGAGAGAGCGCTTTCTGAGCGGCTTCGGTCGTACATGATCCGTCATATTTCTG	3721		
QY	662	TGTTCTGGACAAGATGCCCTCTCAAGCCCAATGGTAAAGTTGACCGGAAGAACT	715		
Db	3722	AACGGTGGATAAATCCGCTTCGGCGAAGCGAAAGTGGACCGAGCGCGCT	3775		

```

RESULT 4
US-09-974-300-2169
; Sequence 2169, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCES: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974, 300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2169
; LENGTH: 8268
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2169

```

Query Match	4.7%;	Score 46.8;	DB 10;	Length 8268;
Best Local Similarity	63.2%;	Prod. No. 8.6e-05;		
Matches	72;	Conservative	0;	Mismatches 42;
			Indels	0;
			Gaps	0;
QY	602	CGAAGTCCGCGAGAGGCTTTGGTCTTACTTCCATCGTACATGATCCCATCGAACATCGT	661	
Db	3363	CGAAATGAGAGAGCGCTTTCTGAGCGGCTTCGTCGTACATGATCCGTCATATTTCGT	3422	
QY	662	TGTTCTGGACAGATGCCCTCTCAACGCCAATGGTAAAGTTGACCGGAAGAACT	715	
Db	3423	AACGGTGGATAAAATCCCGCTTCGGCGCAACGAAAGTCGACGACGCGCGCT	3476	

RESULT 5
US-09-976-059-1
; Sequence 1, Application US/09976059
; Patent No. US20020164747A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin
; FILE REFERENCE: 3019-PCT


```

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7773
; LENGTH: 7347
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(7347)
; US-09-815-242-7773

Query Match          4.0%; Score 40.4; DB 10; Length 7347;
Best Local Similarity 53.9%; Pred.No. 0.013;
Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY      607  TCGCGAGAGCGTTCGGTCCCTTACTTCCATCGTACATGCCATCGAAGCAATCGTTCTTC 666
        ||||| ||| | | | | | | | | | | | | | | | | | | | | | | |
Db      6089  TCGCGGAGTCCGCTGAAGCGGCACCTGCGCGACTACATGTCGCGGCGCACCTGATGCTGC 6148
        ||||| |||| | | | | | | | | | | | | | | | | | | | | | |

QY      667  TGGCAAGAATGCCTCTCAACGCCAATGGTAAAGTTGACCGGGAAGGAACCTCTCTCGCAGGG 726
        ||||| |||| | | | | | | | | | | | | | | | | | | | | | |
Db      6149  TGGAGCGGATGCCGCTCACCGGTCAATGGCAAGCTCGACCGCAGGCGTTCGCCGACCCGG 6208
        ||||| |||| | | | | | | | | | | | | | | | | | | | | | |

QY      727  CAAAGTTGTACCAAGCAGCAGCAGCAGCGCC 760
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      6209  ATCGAGACTTGTGCGCAGCAGCGCCTATTCGAGCGCC 6242
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 7
US-09-974-300-2175
; Sequence 2175, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2175
; LENGTH: 3798
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2175

Query Match      4.0%; Score 39.8; DB 10; Length 3798;
Best Local Similarity 50.7%; Pred. No. 0.015;
Matches 185; Conservative 0; Mismatches 162; Indels 18; Gaps 3;

QY      630  CTTCATCGTACATGATCCATCGAACATCGTTGTCTTGACACAAGATGCTCTCAACGCC 689
      || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      2740  CTACCGGCTTACATGGTCCGGCTGGGTATATCATGCTTGAAGAGCTGCCGTTACCGCT 2799

QY      690  AATGGTAAAGTTGACCGGAAGGAAGTCTCTCGCAGGGCAAGAGTTCTACCGAAGCAGAG 749

```

Db	2800	AACGGAAAAGTCAACCG--ACGCTTCTCCCTGAAGCCGATGGCGCGCGAAATTCACACG	2857
Qy	750	ACAGCAGCGCGTTTACGCACATTTCCCATCAGTCAGTCAGTCGAAGTCATCTTTTGGCAAGAA	809
Db	2858	AGCGCCCTGC-----ACCCCGCAATCTACGGAAGAAAAGCTTGCAAGT	2904
Qy	810	GCCACTGAGGTGTTTGGCA--TGAAGGTTGCACATTACCGATCACTTCTTCAATCTCGGT	866
Db	2905	TGGTCTGAGGTTCTGCGCAGACACGATCGGAATAGACGAACTTTTTTGAATCGGG	2964
Qy	867	GGACACTCTCTCTTTGGCCACGAAGCTCATTTCTCGTATCGACCAACGACTCAAGTCCGT	926
Db	2965	GGACATTCATTGAAAGCCATGGCGCTCGCCTCCCGGATGCTGAAAGATCTCGCAATCGAT	3024
Qy	927	ATCACTGTCAAGGATGCTTTTGACCATCTCTGTATTTCGGGATCTAGCATCTGTCTATCCGT	986
Db	3025	GTTCGGGTGAATGTGTTGTTTGAATGCCACATCGAGGGCCTTCGGGCTTATATCGAC	3084
Qy	987	CAAGG 991	
Db	3085	CAAGG 3089	
RESULT 8			
US-09-974-300-2193			
; Sequence 2193, Application US/09974300			
; Patent No. US2020146721A1			
GENERAL INFORMATION:			
; APPLICANT: Berka, Randy M.			
; APPLICANT: Clausen, Ib Groth			
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene			
; TITLE OF INVENTION: Expression			
; FILE REFERENCE: 10085.500-US			
; CURRENT APPLICATION NUMBER: US/09/974,300			
; CURRENT FILING DATE: 2001-10-05			
; PRIOR APPLICATION NUMBER: 09/680,598			
; PRIOR FILING DATE: 2000-10-06			
; PRIOR APPLICATION NUMBER: 60/279,526			
; PRIOR FILING DATE: 2001-03-27			
; NUMBER OF SEQ ID NOS: 8481			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 2193			
; LENGTH: 3798			
; TYPE: DNA			
; ORGANISM: Bacillus licheniformis			
US-09-974-300-2193			
Query Match 4.0%; Score 39.8; DB 10; Length 3798;			
Best Local Similarity 50.7%; Pred. No. 0.015;			
Matches 185; Conservative 0; Mismatches 162; Indels 18; Gaps 3;			
Qy	630	CTTCCATCTGATGATGCCATCGAATCGTTGTTCTTGACAAAGATGCTCTCAACGCC	689
Db	2740	CTACCGGCTTACATGTTGCGCGCTGGCGTATATCATGCTTGAAGAGCTGCGCGTTACCGCT	2799
Qy	690	AATGGTAAAGTTACCGGAAGAACTCTCTCGCAGGGCAAGGTTGTACCGAAGCAGCAG	749
Db	2800	AACGGAAAAGTCAACCG--ACGCTTCTCCCTGAAGCCGATGGCGCGCGAAATTCACACG	2857
Qy	750	ACAGCAGCGCGTTTACCGACATTTCCCATCAGTCAGTCGAGTCGAAGTCATCTTTTGGCAAGAA	809
Db	2858	AGCGCCCTGC-----ACCCCGCAATGCTACGGAAGAAAAGCTTGCAAGT	2904
Qy	810	GCCACTGAGGTGTTTGGCA--TGAAGGTTGACATTTACCGATCAGTCTTCTCAATCTCGGT	866
Db	2905	TGCTCTGAGGTTCTCGGCAGACAGCAGATCGGAATAGACGAACTTTTTTGAATCGGG	2964
Qy	867	GGACACTCTCTCTTTGGCCACGAAGCTCATTTCTCGTATCGACCAACGACTCAAGTCCGT	926
Db	2965	GGACATTCATTGAAAGCCATGGCGCTCGCCTCCCGGATGCTGAAAGATCTCGCAATCGAT	3024
Qy	927	ATCACTGTCAAGGATGCTTTTGACCATCTCTGTATTTCGGGATCTAGCATCTGTCTATCCGT	986


```
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 234
; LENGTH: 1160
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-123-155-234

Query Match      3.8%; Score 38; DB 9; Length 1160;
Best Local Similarity 9.1%; Pred. No. 0.031;
Matches 68; Conservative 213; Mismatches 470; Indels 0; Gaps 0;

QY 100 CGTATCGATATCGCGTGTGTGTCACGTTCCGGGTTCACCTTGGAGATGACCTTGTGCTTC 159
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1073 S.D.AYS.BA..A.KCTDBDHABDGDK.WAC.NSTBK.BB..AC...HBHHTYRN.CNRY 1014
QY 160 CGTTGAGAAGATGATGCTGATCGACTTTCAAGCAATCAATTGAACCAAGTCACTGG 219
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1013 ..S.TD.D.AD.HBH.CKT.KRWBY.C.RKA.CY.NTR.MCN.TDKMNS...A.H..N. 954
QY 220 GTGACCTTCACAACTTCAGATGCTGCTATCATGAGCAGTACGACAAATTCCTTCGAAA 279
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
953 SC.H.SCDATBN...RWSMN.H.....T..YHKHSYCCD....CM.G.T..CB.CSRKH 894
QY 280 TCACGCGCTTTGAAGACAGGTGCTGCTTCCCTCAATAGCAACATCGATGATGGCAGC 339
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
893 M..HBYSRR.WSBGTSHN.C.CSY...NW.RABB.C..TT..WBAAS...BCDDBSR.YS 834
QY 340 TATCAACCAATTCGGTCCAGCGCGGAGGACATCATCATCTCCGTCCTCCGACATCTTC 399
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
833 TNNM.BBA..HCB.CM...C.C.D...CAY.Y.Y.ACHAR.Y.BTMRYSC..CHMS 774
QY 400 GCATTCGCTGGGAAGCGGGTCCGTGTCAGTCACTGTCGACGACAGTGTCTCAGA 459
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
773 ..RCRSHM.S.NDW.Y.WSYH.GRHW.B..TK.RR.YTT.RY.Y.AT.DC.G.SBNR.T 714
QY 460 ATGGTGCATGGAGCGTGTTCATCATGTTGCTCCCAAGGCGTACTCTGCTCAACT 519
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
713 YKCT.YDSC.T.GR...AC.NH...K.TRB.AD...KTCMYKC...BANC...KECDBHA.A 654
QY 520 TTCTACGGACCATCACCTTCAGAGGTCTGATCTCTCCACCAATCGACCCCTTCAGCGAC 579
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
653 C.C..DWTBAHCKGH.Y.NS...CN.TDKYNH..C.H.S..Y.H.KTIT.NTNSRW 594
QY 580 TCGAAAACCGTGTATCGGCATCGAAGTCCGCGAGAGGTTCGGTCTTACTTCCATCGT 639
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
593 SMN.H.BA.....MHBGM.MCHTC.TCTYGB...CB.CSNAHY...HTTS.RKWA..GT.H 534
QY 640 ACATGATCCCATCGAATCGTGTGTTCTCGAAGATGCTCTCAACGCAATGTTAAAG 699
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
533 .TA.CTD..BAWTRABM...MTB.C.RSSCHNRBAC.RHM.R.BCD..KS...STYNR.B 474
QY 700 TTGACCGGAGGAACCTCTCTCGCAGGGCAAAGTGTACCGAGCAGCAGCAGCAGCGC 759
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
473 B..AHCB.TV.B.C.C.DVH...DKM...A.H..T.RYBMRACC.YSSM...MHS.S 414
QY 760 CGTTACCGACATTTCCCATCAGTGTGAGTCAATCTTTCGGAAGAAGCCATGAGG 819
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
413 C..Y.NMCASCHDK.CRHW...DT...R.YBMCA..HB..TKSGSMBMR.T.KCHY.DS 354
```

```
QY 820 TGTTTGGCATGAGGTTGACATTACCGATCA 850
Db : : : : : : : : : : : : : : : : : : : :
353 NB.GS..WAC..W.BK.TAB.AT...NTBHA 323
```

RESULT 13

```
US-10-184-644-592
; Sequence 592, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 592
; LENGTH: 499
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-592

Query Match      3.7%; Score 37.2; DB 9; Length 499;
Best Local Similarity 9.5%; Pred. No. 0.036;
Matches 36; Conservative 113; Mismatches 228; Indels 0; Gaps 0;

QY 138 CTTGGAGATGAGTGTGCTTCGGTTGAGAAAGATGACTGGATCGACTTTCAGCGCAAT 197
Db : : : : : : : : : : : : : : : : : : : :
9 WLLVLRLPWRVPGQLDPSTGRRFSEHKLCADDECSMMYRGEALDFDTPDCRFVFKKG 68
QY 198 CAATTGAACAGAGTCACTGGTGACCTTCTCAAGTCTTCAGATGCTGCTATCATGCA 257
Db : : : : : : : : : : : : : : : : : : : :
69 DPVYVYKILARGPEWAGSVGRTFGPKDLIOVHEXKEELQVPTDETDFVCFDGR 128
QY 258 GTCAGCAAAATTCCTTCGAAATCAGGCGCTTTGAAAGACAGGTCGCTTCCTCAAT 317
Db : : : : : : : : : : : : : : : : : : : :
129 DDFHNINVELLGLFLELYNSAATDSEKAVEKTLDMEKNPELSKERPEPEPEANSEES 188
QY 318 AGCAACATCATGAGTGGCAGCTATCAACCATTCGGTCCAGCGCGGAGGCGACTCATCA 377
Db : : : : : : : : : : : : : : : : : : : :
189 DSVFSENTEDLQEQFTQKHSHANSOANHAQEQASFESFEEMQLQKLVPESENKTS 248
QY 378 CTATCGCTTCCCGACATCTTCGATGCTGGGGAAGCCGGTTCGGTGTGAGGTCAGT 437
Db : : : : : : : : : : : : : : : : : : : :
249 NSSQVSNQDKIDAYKLLKKEMTDLTKFKFGSTADALVSDDETRLVTSLEDDFDELD 308
QY 438 TCTGCACGACAGTGGTCTCAGAATGTCATGTCAGGCTGTTTTCATCATGTTGTCTCC 497
Db : : : : : : : : : : : : : : : : : : : :
309 EYAVAVKDEENQEDFDELPLLTFTDGEDMKTAKSGVEKYPYTDKEONSNEEDKVQLTVP 368
QY 498 CAAGGCGTACTCTGGT 514
Db : : : : : : : : : : : : : : : : : : : :
369 PGIKNDKNIILTTWGT 385
```

RESULT 14

```
US-10-184-634-592
; Sequence 592, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
```

APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C217
CURRENT APPLICATION NUMBER: US/10/184,634
CURRENT FILING DATE: 2002-06-28
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 592
LENGTH: 499
TYPE: PRT
ORGANISM: Homo Sapien
US-10-184-634-592

Query Match 3.7%; Score 37.2; DB 9; Length 499;
Best Local Similarity 9.5%; Pred. No. 0.036;
Matches 36; Conservative 113; Mismatches 228; Indels 0; Gaps 0;

QY 138 CTGGAGATGAGCTGTGCTTCCGGTTGAGAAAGATGACTGGATCGACTTTCAGCGAAT 197
DB 9 VMLLVLRPWRVPGQLDPSTGRFRSEHKLCADDECSMYRGEALDFTGDCRFVFKKG 68
QY 198 CAATGACAGCAAGTCACTGGTGACCTTCTCAAGTCTTCAGATCGTGTATCATGGCA 257
DB 69 DPVYVYKLRGWPEWVAGSVGRFTGYFPKDLIQVYVHEXTKEELQVPTDTEVCFDGG 128
QY 258 GTCAGCAAAATCCCTTTCGAAATCACGGCTTTGAAAGACAGCTCGCTCCCTCAAT 317
DB 129 DFPHNTNVELGLFLYNSAATDSEKAVEKTLQDMKNPELSKERPEPEPEANSEES 188
QY 318 AGCAATGATGATGGAGCTATCAAGCTATCGGTCAGCGCGGAGGCGACTCATCA 377
DB 189 DSVFSENTEDLOPQFTTKRHSHANSOAHQAQEQASFESFEMLQDLKLVPESENKTS 248
QY 378 CTATCGGTCGCGACATCTTTCGATGCTGGGAGCGGGTCCGCTGTCGAGGTCAGT 437
DB 249 NSSQVSNEQDKIDAYKLLKEMTLDLKTRFGSTADALVSDDETTRLVTSLEDDFDBELDT 308
QY 438 TCTGACGACAGTGGTCTCAGATGCTGACGCTGTTTCCATCATTTGCTCTCC 497
DB 309 EYAVGKDEENEDFELPLLTFTDGEDMKTPAKSGVEKYPTDKBQNSNEEDKVOLTVP 368
QY 498 CAAGGCGTACTCTGTT 514
DB 369 PGIKNDKNLLTWGDT 385

RESULT 15
US-10-123-155-536
Sequence 536, Application US/10123155
Publication No. US20030068794A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria

APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C30
CURRENT APPLICATION NUMBER: US/10/123,155
CURRENT FILING DATE: 2002-04-15
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 536
LENGTH: 513
TYPE: PRT
ORGANISM: Homo Sapien
US-10-123-155-536

Query Match 3.6%; Score 36.4; DB 9; Length 513;
Best Local Similarity 7.2%; Pred. No. 0.068;
Matches 27; Conservative 105; Mismatches 245; Indels 0; Gaps 0;

QY 10 TCTTCACTCGTTGAAAGACAGAGCTTCCAGTCTCTGGTGAACATGTTGAGATCCTGCCAA 69
DB 88 TWLYLDHNHISNIDENAFNGIRKELILSSNRISYFLNNTFRPVTNLRNLDLSYNOLHS 147
QY 70 AGAACATGGAAGCTGTAATGAGCTCAGTCGCTATCGATATGCGGCTGTTGTGCACGCTTC 129
DB 148 LGSEQFRGLRKLKLLSLHLSNSLRTPVRFQDCRNLELLDLGYNRIRSLARNVFAGMRL 207
QY 130 GGGTTCACCTGGAGATGAGCTTGTGCTTCCGGTTGAGAAAGATGAGTGGATCGACTTTC 189
DB 208 KELHLEHNQFSLNLALFPRLVSLQNLVQLQWKNISVIGOTMSWTWSSLQRLDLSGNEIEA 267
QY 190 AAGCAATCAATTGAACAGAGTCACTGGTGGTACCTTCTCAAGTCTTTCAGATCCTGCTCA 249
DB 268 FSPSPVQCPNQLRNLNLSNKLFTTQGEILDWSISLNDISLAGNIWECNRNCSLVNWL 327
QY 250 TCATGGCAGTCAGCAAAATTCCTTTCGAAATCAGCGCTTTTGAAGACAGGTCGCTCGCTT 309
DB 328 KSKGLRENTIICASPKELOGVNVDAVKNVYSIGKSTTERFDLALALPKPTFKPLPRP 387
QY 310 CCCTCAATAGACATCGATGAGTGGCAGCTATCAACCATTCGCTCAGCCGCCGAGGCG 369
DB 388 KHESKPPPLPTVGATEPGETDADAHEISPHKIITAGSVALFSLVLVILLVYVSWKRYPA 447
QY 370 ACTCATCACTATCGGT 386
DB 448 SMKQLOQRSLMRRHKK 464

Search completed: June 4, 2003, 23:54:11
Job time : 173 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2003, 17:53:27 ; Search time 1708.5 Seconds
(without alignments)
9488.839 Million cell updates/sec

Title: US-09-482-788-1_COPY_7000_8000
Perfect score: 1001
Sequence: 1 aacgtgcttttcaactcg.....tcgtcaaggctgggttg 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :			
EST:*			
1:	em_estba:*		
2:	em_esthum:*		
3:	em_estim:*		
4:	em_estmu:*		
5:	em_estov:*		
6:	em_estpl:*		
7:	em_estro:*		
8:	em_hic:*		
9:	gb_est1:*		
10:	gb_est2:*		
11:	gb_hic:*		
12:	gb_est3:*		
13:	gb_est4:*		
14:	gb_est5:*		
15:	em_estfun:*		
16:	em_estom:*		
17:	gb_gss:*		
18:	em_gss_hum:*		
19:	em_gss_inv:*		
20:	em_gss_pin:*		
21:	em_gss_vrt:*		
22:	em_gss_fun:*		
23:	em_gss_mam:*		
24:	em_gss_mus:*		
25:	em_gss_other:*		
26:	em_gss_pro:*		
27:	em_gss_rod:*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	41.2	4.1	744	17	AQ163140
2	40.4	4.0	606	17	AQ990481
3	40.4	4.0	632	17	AQ990701
4	39.8	4.0	850	17	CNS077YR
5	39.8	4.0	1040	17	CNS078CL
6	39.4	3.9	708	17	AQ989957

7	37.6	3.8	619	10	BE492765
8	36.6	3.7	903	17	CNS028CI
9	36.4	3.6	508	9	AL385769
10	36.2	3.6	342	12	BF809723
11	36.2	3.6	353	12	BF094070
12	36.2	3.6	781	13	BM044659
13	36.2	3.6	790	13	BM048043
14	36.2	3.6	792	13	BM044732
15	36.2	3.6	812	12	BG823879
16	36.2	3.6	828	14	BQ900191
17	36.2	3.6	901	14	BQ936677
18	36.2	3.6	949	14	BQ919273
19	36	3.6	628	13	BI180727
20	35.6	3.6	602	14	BQ916759
21	35.6	3.6	812	12	BG319806
22	35.6	3.6	1050	14	BQ277781
23	35.4	3.5	653	13	BI174975
24	35	3.5	360	14	C08494
25	35	3.5	605	10	BB656627
26	35	3.5	638	10	AW740666
27	35	3.5	986	17	CNS00139
28	34.8	3.5	518	9	AL372468
29	34.8	3.5	579	12	BE999591
30	34.8	3.5	579	14	BQ702499
31	34.8	3.5	749	14	BQ655623
32	34.8	3.5	782	17	AQ362655
33	34.6	3.5	330	12	BF745442
34	34.6	3.5	757	9	AL568342
35	34.4	3.4	408	12	BF517011
36	34.4	3.4	437	9	AL377071
37	34.4	3.4	501	14	C18042
38	34.4	3.4	942	17	CNS0102V
39	34.2	3.4	333	13	BI192067
40	34.2	3.4	572	9	AA272108
41	34.2	3.4	686	9	AL581888
42	34.2	3.4	792	10	BE052874
43	34.2	3.4	895	12	BG531334
44	34.2	3.4	1056	17	CNS022J7
45	34	3.4	286	10	BB721296

ALIGNMENTS

RESULT 1
AQ163140
LOCUS mgxb0023A03r CUGI Rice Blast BAC Library Magnaporthe grisea genomic clone mgxb0023A03r, DNA sequence.
DEFINITION AQ163140
ACCESSION AQ163140.1 GI:3559541
VERSION GSS.
KEYWORDS Magnaporthe grisea.
SOURCE Magnaporthe grisea
ORGANISM
REFERENCE 1 (bases 1 to 744)
AUTHORS Yu, Y., Zhu, H., Boyd, C.A., Gaudette, B., Gayle, A., Kingsbury, R., Phillips, K., Sasnowski, M., Wing, R.A. and Dean, R.A.
TITLE A BAC End Sequencing Framework to Sequence the Magnaporthe grisea Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Dean RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson University, Clemson, SC 29634
Tel: 864 656 5737
Fax: 864 656 4293
Email: rdean@clemson.edu
Seq primer: GGAACAGCTATGACCATG
Class: BAC ends
High quality sequence stop: 431.
Location/Qualifiers

```

source      1. 744
/organism="Magnaporthe grisea"
/strain="70-15"
/db_xref="taxon:148305"
/clone="mgxb0023A03r"
/clone_lib="CUGI Rice Blast BAC Library"
/tissue_type="Protoplasts"
/lab_host="E. coli DH10B"
/notes="vector: pBACWICH; Site_1: HindIII; Site_2: HindIII;
Rice blast is one of the most devastating fungal diseases
of rice world wide. It is a filamentous ascomycete with
a haploid genome (n=7) of approximately 40 Mbp. Rice
blast is an important model fungal pathogen for studying
numerous aspects of the fungal-host interaction. In
order to facilitate genome wide analysis, a BAC library
containing 9216 clones with an average insert size of 130
kbp was constructed. This library represents greater
than 25X genome coverage. High density colony filters
are available upon request."
BASE COUNT 169 a 243 c 178 g 154 t
ORIGIN

Query Match      4.1%; Score 41.2; DB 17; Length 744;
Best Local Similarity 56.7%; Pred. No. 0.94;
Matches 76; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 586 ACCGTCGATCGCATCGAAGTCGCGGAGAGGTTGGTCTTACTTCCATCGTACATGA 645
      |||||
Db 71 ATCGTCGATCGAGGTCGAGCTTCGCGACCATCTCAAGTCAAGTTGCCGACCTACGCCG 130
      |||||

QY 646 TCCCATCGAACATCGTTGTTCTGACAAAGATGCTCTCAACGCCAATGGTAAAGTTGACC 705
      |||||
Db 131 TTCTAGCATCTTCATGTCCTGGAGAAGCTGCAATTAACCCCAAGCGGAAGTTGACA 190

QY 706 GGAAGGAACCTCTCT 719
      |||||
Db 191 AGCCCAATCTACCT 204

RESULT 2
AQ990481
LOCUS      606 bp DNA linear GSS 14-AUG-2000
DEFINITION RfC01258 Photorhabdus luminescens strain W14 M13 library
ACCESSION  AQ990481
VERSION     AQ990481.1 GI:9649075
KEYWORDS   GSS.
SOURCE     Photorhabdus luminescens.
ORGANISM   Photorhabdus luminescens.
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Photorhabdus.
REFERENCE  1 (bases 1 to 606).
            ffrench-Constant, R.H., Waterfield, N., Burland, V., Perna, N.T.,
            Daborn, P.J., Bowen, D. and Blattner, F.R.
            A genomic sample sequence of the entomopathogenic bacterium
            Photorhabdus luminescens W14: potential implications for virulence
            Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
JOURNAL    20378633
MEDLINE
COMMENT    Contact: ffrench-Constant RH
            Department of Biology and Biochemistry
            University of Bath
            South Building, Bath BA2 7AY, UK
            Tel: (44) 1225 826621
            Fax: (44) 1225 826779
            Email: bssr@bath.ac.uk
            This is one of 2,122 random reads from the M13 library. For
            annotation of identified clones (BLASTX, BLASTN and mapping to E.
            coli K12 genome) please see ffrench-Constant et al. 2000, Nucleic
            Acids Res.
            Seq primer: M13 Forward
            Class: shotgun.
            Location/Qualifiers
            source      1. .606
            /organism="Photorhabdus luminescens"
            /strain="W14"
            /db_xref="taxon:29488"
            /clone="pLG01258"
            /clone_lib="Photorhabdus luminescens strain W14 M13
            library"
            /dev_stage="primary phase variant"
            /note="Genomic DNA from strain W14 was size selected (1-2
            kb) and then cloned into M13 Janus."
            BASE COUNT 192 a 122 c 115 g 174 t 3 others
            ORIGIN

Query Match      4.0%; Score 40.4; DB 17; Length 606;
Best Local Similarity 53.9%; Pred. No. 1.4;
Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 835 TTGCATTACCGATCATTCTTCAATCTCGGTGACACTCTCTTTGGCCACGAAGCTCA 894
      |||||
Db 268 TTGGTATTTATGATAATTTCTTCACTCGGTGTCATCTCACTTACGAGTTAAATAA 327
      |||||

QY 895 TTTCTCGTATCGACCAACGACTCAAGTCGCTATCACTGTCAAGGATGTCTTTCACCATC 954
      |||||
Db 328 CCCTGCTCTATAAGAAATATGTTAACCAATCATTTCACTATGCATGAGTTGTTCCGAATC 387
      |||||

QY 955 CTGTATTGCGGATCTAGCATCTGTATCCGTCA 988
      |||||
Db 388 CAACCATTTGCTCAGATTGCCAGTGTGATTCGCCCA 421
      |||||

RESULT 3
AQ990701/c
LOCUS      632 bp DNA linear GSS 14-AUG-2000
DEFINITION RfC01513 Photorhabdus luminescens strain W14 M13 library
ACCESSION  AQ990701
VERSION     AQ990701.1 GI:9649295
KEYWORDS   GSS.
SOURCE     Photorhabdus luminescens.
ORGANISM   Photorhabdus luminescens.
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Photorhabdus.
REFERENCE  1 (bases 1 to 632).
            ffrench-Constant, R.H., Waterfield, N., Burland, V., Perna, N.T.,
            Daborn, P.J., Bowen, D. and Blattner, F.R.
            A genomic sample sequence of the entomopathogenic bacterium
            Photorhabdus luminescens W14: potential implications for virulence
            Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
JOURNAL    20378633
MEDLINE
COMMENT    Contact: ffrench-Constant RH
            Department of Biology and Biochemistry
            University of Bath
            South Building, Bath BA2 7AY, UK
            Tel: (44) 1225 826621
            Fax: (44) 1225 826779
            Email: bssr@bath.ac.uk
            This is one of 2,122 random reads from the M13 library. For
            annotation of identified clones (BLASTX, BLASTN and mapping to E.
            coli K12 genome) please see ffrench-Constant et al. 2000, Nucleic
            Acids Res.
            Seq primer: M13 Forward
            Class: shotgun.
            Location/Qualifiers
            source      1. .632
            /organism="Photorhabdus luminescens"
            /strain="W14"
            /db_xref="taxon:29488"
            /clone="pLG01513"
            /clone_lib="Photorhabdus luminescens strain W14 M13
            library"
            /dev_stage="primary phase variant"
            /note="Genomic DNA from strain W14 was size selected (1-2
            kb) and then cloned into M13 Janus."
            BASE COUNT 189 a 116 c 120 g 204 t 3 others
            ORIGIN

```


5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

source
1..1040
/organism="Pichia angusta"
/strain="CBS 4732"
/db_xref="taxon:4905"
/clone="BB0AA019F07"
/clone_lib="BB0AA"
/notes="end : T7"
misc_feature
8..>1027
/note="similar to Saccharomyces cerevisiae ORF YBR115c [LYS2 : L-aminoadipate-semialdehyde dehydrogenase, large subunit 1]"
/evidence="not experimental"
BASE COUNT 256 a 262 c 287 g 234 t 1 others
ORIGIN
Query Match 4.0%; Score 39.8; DB 17; Length 1040;
Best Local Similarity 56.5%; Pred. No. 3.3;
Matches 74; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
QY 852 TTCTTCAATCTCGTGGACACTCTCTTGGCCACGAGCTCATTTCTGATCGACCAA 911
||||| ||| ||||||||||| ||| ||||||| ||| ||||| ||| |||
Db 602 TTCTTGACCTGGGTGGACACTCGATCCTCGCCACGCGGATGATTTTGAGTTGAGAAAA 661
QY 912 CGACTCAAGTCGCTGCTACTGTCAAGATGCTTTGACCATCTGTATTTCGGGATCTA 971
||||| ||| ||||||||||| ||| ||||||| ||| ||||| ||| |||
Db 662 CAGCTGCCGTCGATATCCCTCTCGGAGCCATCTTCAACACCCACCGCTTGCTTCATTC 721
QY 972 GCATCTGTCAT 982
||||| ||| |||
Db 722 GCAGCAGAGAT 732

RESULT 6

AQ989957 708 bp DNA linear GSS 14-AUG-2000
LOCUS
DEFINITION
Photothabdus luminescens strain W14 M13 library
Photothabdus luminescens genomic clone PLG00641, DNA sequence.
ACCESSION
AQ989957
VERSION
AQ989957.1 GI:9648551
KEYWORDS
GSS.
SOURCE
Photothabdus luminescens.
Photothabdus luminescens
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Photothabdus.
REFERENCE
1 (bases 1 to 708)
french-Constant,R.H., Waterfield,N., Burland,V., Perna,N.T.,
Daborin,P.J., Bowen,D. and Blattner,F.R.
A genomic sample sequence of the entomopathogenic bacterium
Photothabdus luminescens w14: potential implications for virulence
Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
20378633
COMMENT
Contact: french-Constant RH
Department of Biology and Biochemistry
University of Bath
South Building, Bath BA2 7AY, UK
Tel: (44) 1225 826621
Fax: (44) 1225 826779
Email: bssrf@bath.ac.uk
This is one of 2,122 random reads from the M13 library. For
annotation of identified clones (BLASTX, BLASTN and mapping to E.
coli K12 genome) please see french-Constant et al. 2000, Nucleic
Acids Res.
Seq primer: M13 Forward
Class: Shotgun.

FEATURES

source
1..708
/organism="Photothabdus luminescens"
/strain="W14"
/db_xref="taxon:29488"

/clone="PLG00641"
/clone_lib="Photothabdus luminescens strain W14 M13
library"
/dev_stage="primary phase variant"
/note="Genomic DNA from strain W14 was size selected (1-2
kb) and then cloned into M13 Janus."
BASE COUNT 220 a 150 c 152 g 183 t 3 others
ORIGIN
Query Match 3.9%; Score 39.4; DB 17; Length 708;
Best Local Similarity 52.8%; Pred. No. 3.1;
Matches 85; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
QY 814 CTGAGGTGTTGGCATGAAGGTTACATTAACCGATCACTTCTTCAATCTCGGTGGACACT 873
||||| ||| ||||||||||| ||| ||||||| ||| ||||| ||| |||
Db 18 CAGTGTGGCGGTAGTCAAGTCGCGATTATGACAACTTTTCGTCCTGGTGGTCACT 77
QY 874 CTCTCTTGGCCAGAGCTCATTTCTGTCGATCGACCAACGACTCAAGGTCGGTATCACTG 933
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78 CTCTTCAAGCAACCGCGCTTATCAGCTTAATAGCAATGATCTAAACATTGAAATACCGT 137
QY 934 TCAGGATGCTTTGACCATCCTCTATTTCGGGATCTAGCA 974
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 138 TGACGACTGCTTTGAACATCAACACTTGAGCGGATTAGCA 178

RESULT 7

BE492765 619 bp mRNA linear EST 16-APR-2001
LOCUS
DEFINITION
WHE0564_E04_E04ZE Triticum monococcum vegetative apex cDNA library
Triticum monococcum cDNA clone WHE0564_E04_E04, mRNA sequence.
ACCESSION
BE492765
VERSION
BE492765.1 GI:9659358
KEYWORDS
EST.
SOURCE
Triticum monococcum.
ORGANISM
Triticum monococcum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae
; Triticeae; Triticum.
REFERENCE
1 (bases 1 to 619)
AUTHORS
Anderson,O.D., Chao,S., Dubcovsky,J., Echenique,V., Han,P.S., Hsia
C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L.,
Stamove,B. and Tong,J.C.
The structure and function of the expressed portion of the wheat
genomes - Vegetative apex cDNA library from Triticum monococcum
Unpublished (2001)
COMMENT
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: andersn@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Stratagene T3 primer.

FEATURES

source
1..619
Location/Qualifiers
/organism="Triticum monococcum"
/cultivar="DV92"
/db_xref="taxon:4568"
/clone="WHE0564_E04_E04"
/clone_lib="Triticum monococcum vegetative apex cDNA
library"
/tissue_type="Vegetative shoot apex"
/dev_stage="Three weeks-old plants"
/lab_host="E. coli XL0LR"
/note="Vector: Lambda pBK-CMV (Lambda Zap Express),
excised phagemid; Site 1: EcoRI; Site 2: XhoI; The tissue,
total RNA, and poly(A) RNA were prepared, a cDNA library
was made, and the cDNA clones were in vivo excised at the
University of California, Davis (V. Echenique, B. Stamova
, J. Dubcovsky). Plasmid DNA preparations and DNA

Qy	858	AATCTCGGTGGACACTCTCTCTTGGCCAGCAAGCTCATTTCTCGTATCGACCAACGACTC		917
Db	139	ACTGTGGGTGACACCCGCTCATGTGGGAATCAGCACATCTTTGACCTCAGCCCAAGAGCGC		80
Qy	918	AAGTCCGTTATCACTGTCRAAGGATG		942
Db	79	AGAGTCCAGGTCCCTTAGCTGGATG		55
RESULT 13				
LOCUS	BM048043/c			
DEFINITION	603620369F1 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:5446001 5', mRNA linear EST 07-NOV-2001			
ACCESSION	BM048043			
VERSION				
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 790)			
TITLE	NIH-MGC http://mgc.nci.nih.gov/			
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)			
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabs-r@mail.nih.gov Tissue Procurement: DCTD/DTP cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLCMI927 row: j column: 18 High quality sequence stop: 781.			
FEATURES				
source				
1..790				
/organism="Homo sapiens"				
/db_xref="taxon:9606"				
/clone="IMAGE:5446001"				
/clone_lib="NIH_MGC_40"				
/lab_host="DH10B (phage-resistant)"				
/tissue_type="carcinoma, cell line"				
/note="Organ: prostate; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."				
BASE COUNT	166 a	223 c	222 g	179 t
ORIGIN				
Query Match	3.6%; Score 36.2; DB 13; Length 790;			
Best Local	53.1%; Pred. No. 31;			
Matches	77; Conservative	0; Mismatches	68; Indels	0; Gaps
Qy	798	CTTTGGCAAGACCCACTGAGGTGTTTGGCATGAAGTTGACATTAACCGATCACTCTTC		857
Db	200	CTGCTCTATGTTTCTCCGTGGTAGACAACCTCCAAAGTGGCCTTTGCCAATGACTCGGTC		141
Qy	858	AATCTCGGTGGACACTCTCTCTTGGCCAGCAAGCTCATTTCTCGTATCGACCAACGACTC		917
Db	140	ACTGTGGGTGACACCCGCTCATGTGGGAATCAGCACATCTTTGACCTCAGCCCAAGAGCGC		81
Qy	918	AAGTCCGTTATCACTGTCRAAGGATG		942
Db	80	AGAGTCCAGGTCCCTTAGCTGGATG		56
RESULT 14				

BM044732/c

LOCUS 603622473F1 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:5447730 5', linear mRNA 792 bp EST 07-NOV-2001
 DEFINITION mRNA sequence.
 ACCESSION BM044732
 VERSION BM044732.1 GI:16773999
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 AUTHORS 1 (bases 1 to 792)
 TITLE NIH-MGC http://mgc.nci.nih.gov/.
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC).
 COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov

Tissue Procurement: DCTD/DP
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCMI932 row: b column: 19
 High quality sequence stop: 786.

FEATURES

source
 1..792
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5447730"
 /clone_lib="NIH_MGC_40"
 /tissue_type="carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: prostate; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
 BASE COUNT 166 a 225 c 224 g 177 t
 ORIGIN

Query Match 3.6%; Score 36.2; DB 13; Length 792;
 Best Local Similarity 53.1%; Pred. No. 31;
 Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
 QY 798 CTTTGGCAAGACCCACTGAGGTGTTGGCATGAAGTTGACATTACCGATCACTTCTTC 857
 DB 200 CTGGTCTATGTAATCTCCGGTGTAGACAACTCCAAAGTGCCCTTCCCAATGACTCGGC 141
 QY 858 AATCTCGGTGGACACTCTCTTGGCCACGAGCTCATTTCTCGTATCGACCAACGACTC 917
 DB 140 ACTGTGGGTGACCAACCGCTCATGGGGAATCAGCACATCTTTGACCTCAGCCAAAGAGCG 81
 QY 918 AAGTCCCGTCACTGCTCAAGGATG 942
 DB 80 AGAGTCCAGGTCCTTAGCTGGATG 56

RESULT 15

BG823879/c
 LOCUS 602729553F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4878444 5', linear mRNA 812 bp EST 22-MAY-2001
 DEFINITION mRNA sequence.
 ACCESSION BG823879
 VERSION BG823879.1 GI:14171466
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 812)
 NIH-MGC http://mgc.nci.nih.gov/.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC).
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: NIH Intramural Sequencing Center
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCMI764 row: j column: 13
 High quality sequence stop: 809.

FEATURES

source
 1..812
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4878444"
 /clone_lib="NIH_MGC_15"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 163 a 235 c 235 g 179 t
 ORIGIN

Query Match 3.6%; Score 36.2; DB 12; Length 812;
 Best Local Similarity 53.1%; Pred. No. 32;
 Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
 QY 798 CTTTGGCAAGACCCACTGAGGTGTTGGCATGAAGTTGACATTACCGATCACTTCTTC 857
 DB 374 CTGGTCTATGTAATCTCCGGTGTAGACAACTCCAAAGTGCCCTTCCCAATGACTCGGC 315
 QY 858 AATCTCGGTGGACACTCTCTTGGCCACGAGCTCATTTCTCGTATCGACCAACGACTC 917
 DB 314 ACTGTGGGTGACCAACCGCTCATGGGGAATCAGCACATCTTTGACCTCAGCCAAAGAGCG 255
 QY 918 AAGTCCCGTCACTGCTCAAGGATG 942
 DB 254 AGAGTCCAGGTCCTTAGCTGGATG 230

Search completed: June 4, 2003, 21:48:28
 Job time : 1715.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2003, 12:33:15 ; Search time 115 Seconds
(without alignments)
3625.576 Million cell updates/sec

Title: US-09-482-788-2

Perfect score: 16128

Sequence: 1 MEYLVAVDGRDLPTTAPSF.....RVEHLLVEVSKTFEGLNSSL 3129

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16128	100.0	3129	21 AAB07427	Amino acid sequenc
2	8799	54.6	3210	22 AAB73958	Mycelia sterilia c
3	5000	31.0	15281	15 AAR44929	T. niveum Cyclosp
4	2146.5	13.3	2841	21 AAB07581	Protein encoded by
5	1785.5	11.1	3587	14 AAR34713	Bacillus subtilis
6	1713.5	10.6	2448	22 AAU36277	Pseudomonas aerugi
7	1677	10.4	4342	22 AAU33611	Pseudomonas aerugi
8	1632	10.1	4999	23 AAQ22159	Ramoplanin biosynt
9	1587.5	9.8	3588	14 AAR34712	Bacillus subtilis
10	1572	9.7	1537	22 AAB83971	Amino acid sequenc

11	1516	9.4	3639	14 AAR40227	ACVS. Acremonium
12	1515	9.4	3712	12 AAR13896	ACV synthetase. A
13	1429	8.9	2323	22 AAG81833	S. epidermidis ope
14	1425.5	8.8	2404	23 ABP38619	Staphylococcus epi
15	1413.5	8.8	3722	12 AAR10145	Cephalosporin anti
16	1406	8.7	2397	22 AAU36672	Staphylococcus aur
17	1375.5	8.5	4999	23 AAQ22158	Ramoplanin biosynt
18	1374	8.5	1668	22 AAU34158	Staphylococcus aur
19	1335	8.3	3778	12 AAR13895	ACV synthetase. P
20	1332.5	8.3	3768	12 AAR13753	ACVS. Penicillium
21	1298	8.0	2162	21 AAB07560	Protein encoded by
22	1048.5	6.5	1591	22 ABB52772	Escherichia coli p
23	1006.5	6.2	2766	22 AAB83972	Amino acid sequenc
24	968	6.0	2675	21 AAB07564	Protein encoded by
25	914.5	5.7	1218	21 AAB07563	Protein encoded by
26	876	5.4	2626	21 AAB07569	Protein encoded by
27	791	4.9	1066	21 AAB07561	Protein encoded by
28	782.5	4.9	1410	21 AAY58574	Sorangium cellulos
29	759	4.7	2392	21 AAB07565	Amino acid sequenc
30	756	4.7	1274	14 AAR34714	Bacillus subtilis
31	684.5	4.2	766	22 AAG81663	S. epidermidis ope
32	675	4.2	1874	22 ABB52779	Escherichia coli p
33	641	4.0	1435	22 ABB52769	Escherichia coli p
34	623	3.9	1051	23 AAQ22157	Ramoplanin biosynt
35	582.5	3.6	760	22 ABB52770	Escherichia coli p
36	571.5	3.5	1103	22 AAU52523	Propionibacterium
37	559	3.5	1293	22 AAU34486	E. coli cellular p
38	554	3.4	1455	22 ABB52766	Escherichia coli p
39	538.5	3.3	1294	22 AAU38131	Salmonella typhi c
40	507	3.1	450	23 AAE14387	Human AMP-binding
41	490	3.0	578	21 AAB07587	Protein encoded by
42	475.5	2.9	891	23 AAQ22162	Ramoplanin biosynt
43	434.5	2.7	1391	21 AAY44644	Candida albicans a
44	424	2.6	1274	22 AAB79202	Corynebacterium ql
45	424	2.6	1295	22 AAG92726	C glutamicum prote

ALIGNMENTS

RESULT 1

AAB07427

ID AAB07427 standard; Protein; 3129 AA.

XX AAB07427;

XX AC

XX 20-OCT-2000 (first entry)

DT

XX Amino acid sequence of a cyclohexadepsipeptide synthetase.

DE

XX Cyclohexadepsipeptide synthetase; filamentous fungal cell;

XX KW cyclohexadepsipeptide; antibiotic.

XX OS Fusarium venenatum.

XX XX

XX WO200042203-A2.

XX PD

XX 20-JUL-2000.

XX XX

XX 13-JAN-2000; 2000WO-US00913.

XX XX

XX 13-JAN-1999; 99US-0229862.

XX XX

XX (NOVO) NOVO NORDISK BIOTECH INC.

XX XX

XX Berka RM, Rey MW, Yoder WT;

XX XX

XX WPI; 2000-482833/42.

XX N-PSDB; AAA58762.

XX XX

XX Producing a heterologous polypeptide for production of antibiotics

XX PT comprises cultivating a mutant of a parent filamentous fungal cell

XX PT comprising a nucleic acid sequence encoding cyclohexadepsipeptide -

Db 1921 GYSKALDENRFVHTVNDQTVKAYRTGDRYRIRIGDGLIEFFGRMDTQFIRGNKRIESA 1980
QY EIEAALLRDSSVRAAAVVLQNEQDAPEILGFVADHDHSENKQGSANQVEGMDHFEES 2040
Db EIEAALLRDSSVRAAAVVLQNEQDAPEILGFVADHDHSENKQGSANQVEGMDHFEES 2040
QY GMYSDIGIDPSTIGSDFKGTSMYDGSQIDFDEMHEWLGETTRTLHDNRSLGNVLEIGT 2100
Db GMYSDIGIDPSTIGSDFKGTSMYDGSQIDFDEMHEWLGETTRTLHDNRSLGNVLEIGT 2100
QY GSGMILFNLDSRLSEYVGLERSRNSAAAFVNKATESIPSLAGKAKVQVGTATDIGOVDLH 2160
Db GSGMILFNLDSRLSEYVGLERSRNSAAAFVNKATESIPSLAGKAKVQVGTATDIGOVDLH 2160
QY PDLVVLNSVIOYFSSSEYLAELADTLIHLPNVQRIFFGDRSQATNEHFLAARAIHTLCK 2220
Db PDLVVLNSVIOYFSSSEYLAELADTLIHLPNVQRIFFGDRSQATNEHFLAARAIHTLCK 2220
QY NATKDDVRQKMAELEDMEELLVEPAFTSLKDRPGLVHVEILPKMNEAVNELSAYRY 2280
Db NATKDDVRQKMAELEDMEELLVEPAFTSLKDRPGLVHVEILPKMNEAVNELSAYRY 2280
QY AAVHVHRSGLDELVLPEKDDWIDFOANQLNOKSLGDLKSSDAAINAVSKIPPEITAF 2340
Db AAVHVHRSGLDELVLPEKDDWIDFOANQLNOKSLGDLKSSDAAINAVSKIPPEITAF 2340
QY ERQVVASLNSNIDWOLSTIRSSAEGDSSLVDPDIFRIAGEAGFRVEYSSAROWSONGAL 2400
Db ERQVVASLNSNIDWOLSTIRSSAEGDSSLVDPDIFRIAGEAGFRVEYSSAROWSONGAL 2400
QY DAVFHCCSQGRTLVNFTDHLRGSLLTNRPQLQNRRIAEVRELRSLSLPSYMP 2460
Db DAVFHCCSQGRTLVNFTDHLRGSLLTNRPQLQNRRIAEVRELRSLSLPSYMP 2460
QY SNIVLDKMPNLNANGKVDKELSRKRAKVVVQKQTAAPLTPPISEVEVILCEEATEVFGM 2520
Db SNIVLDKMPNLNANGKVDKELSRKRAKVVVQKQTAAPLTPPISEVEVILCEEATEVFGM 2520
QY KVDITDHFNFGLGSHSLATKILSRIDQRLKVRITVKDVFDPVADLASVIRQGLGLOOP 2580
Db KVDITDHFNFGLGSHSLATKILSRIDQRLKVRITVKDVFDPVADLASVIRQGLGLOOP 2580
QY VSDGQGQDSRAHMAPRTETAILCDEFAKVLGVQVITDNFDFLGGHSLMATKLAIRIGH 2640
Db VSDGQGQDSRAHMAPRTETAILCDEFAKVLGVQVITDNFDFLGGHSLMATKLAIRIGH 2640
QY RLDTTVSKVDVDFHPVLPQALALDNLVQSKTNEIVGGRMAEYSPFQLLTEDPEEFMA 2700
Db RLDTTVSKVDVDFHPVLPQALALDNLVQSKTNEIVGGRMAEYSPFQLLTEDPEEFMA 2700
QY SEIKPOLBELQEIQDIYESTOMQKAFLEDHTTARPPVPVFFIDFPSTSEPDAAGLIKAC 2760
Db SEIKPOLBELQEIQDIYESTOMQKAFLEDHTTARPPVPVFFIDFPSTSEPDAAGLIKAC 2760
QY ESLVNLHDFIRTVFAEASGELYQVVLSCDLPLQVETEDNTINTATNEFLDEFAKEPVRL 2820
Db ESLVNLHDFIRTVFAEASGELYQVVLSCDLPLQVETEDNTINTATNEFLDEFAKEPVRL 2820
QY GHPLIRFTIINQKTSRMVIMRISHALYDGLSLEHVVRKLHMLYNGRSLLPHPQFSRYQWY 2880
Db GHPLIRFTIINQKTSRMVIMRISHALYDGLSLEHVVRKLHMLYNGRSLLPHPQFSRYQWY 2880
QY TADGREGHGFWRDVIQNTPTILSDDTVWGDNDATCKALHLSKLVNIPSOVLGRSSNII 2940
Db TADGREGHGFWRDVIQNTPTILSDDTVWGDNDATCKALHLSKLVNIPSOVLGRSSNII 2940
QY TOATVFNACALVLSRESKDWFCRTVSGRQGLPVEYQDITVGCTNAVPVRAHIESSD 3000
Db TOATVFNACALVLSRESKDWFCRTVSGRQGLPVEYQDITVGCTNAVPVRAHIESSD 3000
QY YNQLLHDIDQYLLSLPHETIGFSDLKRNCTDWPDAITNFSCCITYHNFEYHPESQFEQ 3060
|||||

Db 3001 YNQLLHDIDQYLLSLPHETIGFSDLKRNCTDWPDAITNFSCCITYHNFEYHPESQFEQ 3060
QY 3061 RVEMGVLTAKFVNIEMDEPLYLDAIAGEVEPDGAGLKVTVIKTQLFGRRKRVEHLLLEEVSK 3120
Db 3061 RVEMGVLTAKFVNIEMDEPLYLDAIAGEVEPDGAGLKVTVIKTQLFGRRKRVEHLLLEEVSK 3120
QY 3121 TFEGLNSSL 3129
Db 3121 TFEGLNSSL 3129
RESULT 2
AAB73958
ID AAB73958 standard; Protein; 3210 AA.
XX AAB73958;
XX AC AAB73958;
XX DT 29-MAY-2001 (first entry)
XX Mycelia sterilia cyclic depsipeptide synthase.
DE Mycelia sterilia; cyclic depsipeptide synthase; anthelmintic;
KW PF1022.
XX Mycelia sterilia.
OS WO200118179-A1.
XX 15-MAR-2001.
PD 07-SEP-2000; 2000WO-JP06103.
XX 07-SEP-1999; 99JP-0253040.
PR 06-APR-2000; 2000JP-0104291.
XX (MEIJ) MEIJI SEIKA KAISHA LTD.
PA Midoh N, Okakura K, Miyamoto K, Watanabe M, Yanai K, Yasutake T;
PI Aihara S, Futamura T, Kleinkauf H, Murakami T;
XX WPI: 2001-265970/27.
DR N-PSDB; AAF79702.
XX Novel cyclic depsipeptide synthase and gene encoding it for efficient
production of anthelmintic substance PF1022
PS Claim 1; Page 59-80; 92pp; Japanese.
XX The present sequence is a cyclic depsipeptide synthase from Mycelia
sterilia. The present protein, or a protein containing an addition,
deletion and/or substitution of one or more amino acid residues is useful
for the efficient production of the anthelmintic PF1022 (cyclo(D-lactyl-
L-N-methylleucyl-D-3-phenyllactyl-L-N-methylleucyl-D-lactyl-L-N-
methylleucyl-D-3-phenyllactyl-L-N-methylleucyl)).
XX Sequence 3210 AA;
Query Match 54.6%; Score 8799; DB 22; Length 3210;
Best Local Similarity 54.6%; Pred. No. 0;
Matches 1768; Conservative 512; Mismatches 777; Indels 184; Gaps 44;
QY 14 PPTPASFCSHGSDPLNSSYEQFLHYGLDSSRIEAIKPTCTPQDMLDNCALDKOSAIGH 73
Db 22 PLLPHDDEKHS- ---NLVEQATRHFGLSRDKIENVLPTCTSEQCDVIDCAVDDRRAIGH 77
QY 74 AYVDVPTDIDISRFALANKEIVNQTPALRAFAFTSDSGTSQVILKDSFVSNMCHSSSS 133
Db 78 VYIDPNTVDIQRLAANKEVVRQTPILRTGTFSETGDSFQIVLKEG-CLPMMYATCLG 136
QY 134 SPDEVVRDEAAAAAGSPRCNREVLLEDMTQKCOLVWTFESHALVDVTFQOVRVLSRFAAY 193
Db 137 MKGAVIQDEAAVAMTGPCRNRVYVLEDPSKORLLIWTESHALVDVTFQOVRVLSRFAAY 196
|||||

QY 194 KHEKTHREPTESSDATDTSQSVSVSMSCEDNAVSAATHFWOHLNDLNASVPHLSL 253
 Db 197 ---DGRDECPRK-----DTEHVS-----RWQOHFEGELDASVFPPLPS 233
 QY 254 HLMVNPPTTAEHRTF--PLSQKALSNAICRTALSILLSRYTHSDEALFCAYTEQS-L 310
 Db 234 HLTVCNPNARAHEHLSYGPV-QRKWSHTSICRAALAVLLSRFTSHSSEALGVVTEOSH 292
 QY 311 PFDKHYLDAGTYOTVAPLRVHCQSNLRASDWDALSSVDDRLGHLAPGLRDRNTGONG 370
 Db 293 SEDQRRSIDGPARTVPIRLVLCAPQYVSDVIGAITAHEHAMRGFEHAGLRNIRRTGGG 352
 QY 371 SAACDFQTVLLVTDGSHNNINGFLOQITTESSHMPCCNRRALLLHCOMESSGALLVAY 430
 Db 353 SAACGFQTVLLVTDGAPKTGCVSLHRSVESDRMPCCANRALLDDCOMAGNSASLVARY 412
 QY 431 DHNVDSLOTRRLQOFGHLIKLOSPLDSSMAEVLNMTEDYRAEIESMNSQPLEVQDT 490
 Db 413 DHNVDPROMSREFRLQLGLYLQFHHVVDLPVRELVDVTAEDCAEIEKWSERLTWQDA 472
 QY 491 LIHHEMLKAVSHSPTKTAIQAWDGDWYSELNDVSSRLAVHIKSLGRAQAQAIIPVFEK 550
 Db 473 LIHDTISKWAAGDNPKAFAVWDGWTYAEJDNLSRLAVYIQSLDLRPGQAILPLCFEK 532
 QY 551 SKWVIASMLAVKSGNAFTLIDPNDPPARTAQVVTQTRATVALTSKLHRETVOKLVGRCV 610
 Db 533 SKWVVATILAVLKVGRAFTLIDPCDPSARMAQVCOQTSAITVALTSKLHNTILRSVSRCI 592
 QY 611 VVDELLQSVASDD-FSSLTKSODLAVIPTSSTGDPKMGIMTEHRAFSKALKFGASL 669
 Db 593 VVDDLLRLSLFADGRILKATVKPQDLAYVIFTSGTGPKEGIMTEHGRFVSCAMKFGPAL 652
 QY 670 GINSDTRALQFTHAFGACILLEIMTTLINGCVCIPISDDDRMNSIPFINRYNNWMMAT 729
 Db 653 GMDHTRALQFASVAFACILVEVYVTLMHGVCVCIPSDDDRLNNVPEFIKRAQVNWVILT 712
 QY 730 PSYNGTSPEDVPGLATLVLGEOMSSVNAIWAPKQLLNGYQSSSSSTCF---ASN 786
 Db 713 PSYIGTQPEDEVPGQLTVLVGEPISASIRDWASQVRLNNAVQSSSTMCSTVYSPL 772
 QY 787 STEPNNNGRAVGAHSWIDPNDINRLVPIGAVGELVITESGIARDYIVPPPPKSPFFTD 846
 Db 773 SLEPNNIGRAVGARSWIIDPEDEPLAPICIGELVITESGIARDYIIPAPKSPPLLA 832
 QY 847 IPSWYPANTPPDGAKYRTGDLARYASDGSITVCLGRIDSQVKRGORVELGALETHLRQ 906
 Db 833 PPAPYPAGKLSNAFKYKTGDLVRYGPDGTIVCLGRKDSQVKRGORVELSAVEASLURQ 892
 QY 907 MPDOLTIVBATKRSOSANSTSLIAFLIGSSYFCN-----RPSDAHILDHDATKAINIK 960
 Db 893 LPSDIMPVAEAIKRSOSSGTVLTAFILIGSSKSGDGNHLSAADAVIDLHDGATNEINAK 952
 QY 961 LEQVLPKHSPTSPYICMLPRATGKIDRRRLRINGKDLIDKQTOGAIVQ-----QAPA 1015
 Db 953 LOQILPOHSPVSYIIHNEILPRATGKADRKRLRSIASKLLGELSNVTSQPTKEKHDA 1012
 QY 1016 PIPVFAADTAALKHSIWQSLGIDPATVNVGATFEELGGNSITAKVNNARSVMGDLKVS 1075
 Db 1013 -----TGIEVKLELWFLSLNLPNSQDVGSFEDLGGNSIIAKVNNARSAGIALKVS 1067
 QY 1076 NIYQHTPLAGISAVVKGDPDLSYTLIPKSTHEGVEQSYSGRLWFLDQLDVGLWLYLIPY 1135
 Db 1068 DIFQNPETLAGLVNDVIGRDPAPYNIPTTAYSGPVEQSFAGRLWFLDQELDALWLLPY 1127
 QY 1136 AVMRGPVNVDAALRALAALQORHETLRTTFEDQGVGVQIVVHEKUSEMKVITDLGCS-D 1194
 Db 1128 AVMRGPVNVDAALRALAALQORHETLRTTFEDQGVGVQIVVHASPISDLRIIDVSGDRN 1187
 QY 1195 LDPFEVINOEQTPPFLNLSSEAGWRATLLRLGEDDHILTIYMHIIISDGWSIDVLRDLNQ 1254
 Db 1188 SDYLOLLHQEQTPPFLACQAGRWVSLIRLGEDDHILSIYMHIIISDGWSIDVLRDLNQ 1247
 QY 1255 LYSAAKLDKSDPLSALTPLPIQYSDFAKWQK--DQFIEQEKOLNYWKOLKSDSPAKIPT 1312

Db 1248 FYSAAIRGS-DPLSVSVPLPHYRDFSVWQKQVEQETHEHEROLEYVWVQLADSSAAEFLLT 1306
 QY 1313 DFARPALLSGDAGCVHVTIDGELYSLAFACNEHNTTSFVVLAAAFRAHAYELTAVEDAV 1372
 Db 1307 DFPRNILSGAGSVPVVIEGELYELQEFCKVEQMTFPAVLGAFRAHYELTGAEDSI 1366
 QY 1373 IGTPIANRNRELEDIIGCFVNTQCMRINIDHDTFTGLINQKATTTAAAFENEDIPPER 1432
 Db 1367 IGTPIANRNRELENMIGFVNTQCMRITVDGDDTFESLVRQVRTTATAAEFHODVPER 1426
 QY 1433 VVSALQPSRDLSTPTLAQLIFAVHSQKDLGRFKQGLSVSPVPSKATRTDMEPHLFOE 1492
 Db 1427 VVTALLPRSDLSRNLPLAQLTFALHSQODLGFLEGLVABFVSNKVYTRFEDVEHLEQ 1486
 QY 1493 TDSLKGSVNFADLFCMETVENVRPFEILNRNGLOSSRTPVSLPILPDGIVTLEKLOVL 1552
 Db 1487 AGRLSGNVFAADLFCPETISNVVAIFQILRQGRIRQPTPIAVLPDGLDGLADRAMGELL 1546
 QY 1553 NVKHVDYPRSSSLADVFQTSAYPDSLVAVDSSCRITYTTELDRQSDILAGMLRRSRMPA 1612
 Db 1547 EIEKAEYPRSSVVDVFRKQVAAPHAFVAVDSASRLTYADLDRSQDQATLWGLRRNTA 1606
 QY 1613 ETLVAVAPRSCETIVAFGVLKANLAYLPDVRSPSARVODIISGLSGPTIVLIGHDTA 1672
 Db 1607 ETLGVILAPRSCQVWAILGILKANLAYLPDVNCPTARLQTLILNHRHLVLLGSNAT 1666
 QY 1673 PPDIEVNTVEFVIRDALND-----SNADGF----- 1698
 Db 1667 TPDVQIPDVELVRISDILDRPINOAKLNGHTKSNYSKPNGYTHLKGSYNLNGYSKONG 1726
 QY 1699 -----EVTEHD-----STKPSATSLAVVLYTSGSTGRPKGVMIHRV 1735
 Db 1727 YAQLNHRERNYLDLNGHLLNSDITTSGPSATSLAVVIFTSGTGRPKGVMIHRV 1786
 QY 1736 IIRTVTSG-CIPNVPSETRAHMATIAFDGASYEIIYALLGRITLVCVDTMTLIDARALK 1794
 Db 1787 IIRLAKNRIISRPSPVAKVAHLSNIAFDAATWEMFAALLNGGLFLVCIDVMTTLIDSTKLE 1846
 QY 1795 DVFFREHVNAASHVTSQSDVPLVRPRLSRTLLMFFFLVVTDS--TAPDALDAQGLYQGV 1852
 Db 1847 AAFAREQINAAITPALLKQCLANIPITLGR---LSALVIGGDRLDGODATAAHALV-GA 1902
 QY 1853 QCVNGYPTENGVMSTIYPIIDSTESFINGVPIGRALNNSGAYVVDPEQOLVIGVMGELY 1912
 Db 1903 GVTNAVYPTENGVIITNTKNDSTINGVPIGCAISNGAYITDPDQQLVPPGVMGELY 1962
 QY 1913 VTGDGLARGYSDKALDENRFRVHTVNDQTVKAYRTGDRVRYRGDGLIEFFGRMDTQFKI 1972
 Db 1963 VTGDGLARGYTDPALDAGRFVQIMINDKAVRAYRTGDRARYRGDGLIEFFGRMDQVKI 2022
 QY 1973 RGNRIESAETEAALLRDSSVYRDAAVVLLQONEDQAPILGFWADHDHSENDKQGSANOVE 2032
 Db 2023 RGRHIEPAEVERAILDQDSARDAAVYVIRHQEGEPEMVGAVTHGDHSA-EQEAADDOVE 2081
 QY 2033 CWQDHFEESGMYSDIGEIDPSTIGSKGTSMDGSDIDFDEHMEWJGETTRTLHDNRSL 2092
 Db 2082 GWKDFPESNTYADMTIGQSAIGNDTGWTSMYDGSSEINKAEMQEWLDDTMTLLDGOAP 2141
 QY 2093 GNVLEITGSGMTILNLDLSRLSVGLPESRSAAAFVNKATSTPSILAGKAKVQVGTATD 2152
 Db 2142 GHVLEITGSGMVLNGLAGLSYVGLPESRSATFTVKAINSTPALAGKAEVHVGTATD 2201
 QY 2153 IGQVDDLHPLVNLNSVIOYFPPSEYLAETADTLIHLPNVQRIFFGDRSQATNEHFLAA 2212
 Db 2202 INRLRLPDLVNLNSVVOYFPTPEYLVVESLVRIPGVKRVFGDIRSHATNRHFLAA 2261
 QY 2213 RALHTLIGKATKDDVRQKMAELEDMEELLVPEAPFTS-LKDRPGLVHEVHEILPKNMEA 2271
 Db 2262 RALHSLGKATKADIRQKMTMEMEREELVDPAPFTALLOGLADRLKIVHEILPKNMEA 2321
 QY 2272 VNELSAYRAAVVHVHVG-SLGDDELVLPEKDDWIDFQANOLNOKSLGDLK-KSSDAALMA 2329

Db 2322 TNELSARYTAVIHVGRPEEQSRVYPIQVNDWIDFQASRIDRRALLRLQSRADAATVA 2381
 QY 2330 VSKIFEITAFERQVVASLNSIDE-----WOLSTIRSSAEGDSSLSVPDIFRIA 2379
 Db 2382 VSNIPYSKTIVERHVHVESLDDNNRTHRAPDGAAW-ISAVESKAERCTSLSTVDLVQLG 2440
 QY 2380 GEAGERVEVSSARQWONGALDAVPHH-----CCSOGRTLVNPPPT-DHHLRGSDDLNTNRL 2434
 Db 2441 EAGERVEVSSARQWONGALDAVPHH-----CCSOGRTLVNPPPT-DHHLRGSDDLNTNRL 2500
 QY 2435 ORLQNRRTAIEVRRLRSLPSMIPSNIVLDMKPLNANGKVRKELSRRAKVVPKQOT 2494
 Db 2501 QRLQSRERASQIREQKAVLPSMIPSLRIVIQMPLNANGKVRKELSRRAKVVPKQOT 2560
 QY 2495 AAPLPTFFPISEV-----EVILCEBATEVFGMKVDITDFFHNLGSHSLATKLISRIDQR 2548
 Db 2561 A-----PAKPVKQVDPVNLAEILCEBATEVFGMKVDITDFFHNLGSHSLATKLISRIDQR 2617
 QY 2549 LKVRITVKDVPDHPVADLASVIROGLGLOQ---PVSDGQ---QDRSAHMAPRTTEAIL 2603
 Db 2618 LKVRITVKDVPDHPVADLASVIROGLGLOQ---PVSDGQ---QDRSAHMAPRTTEAIL 2677
 QY 2604 CDEFKVLGFGVIGITDNFDFLGGHSLMATKLAIRIGHRLDTTVSVKDVDFHPVLPQLATA 2663
 Db 2678 CKEFADVLGVESATDSFDFLGGHSLMATKLAIRIGHRLDTTVSVKDVDFHPVLPQLATA 2737
 QY 2664 LDNLVQSKTNEIVGGRMAEYSPFOLLETPDEEFMAEIKPQLEIQ-EIIDIYPTQM 2722
 Db 2738 I-RLTQAKGHEATNGVQIANDAPFOLISVDEPFIQVREIAPOLQCSPEITLDVYPATQM 2796
 QY 2733 QKAFLEFHTHTARPREVFYIDPSTSEPDAGLTKACESLVNHLIDIFRTVFAEASGELY 2782
 Db 2797 QRVFLNVPVTKGRSPPTFHIDFP--PDADCASLMRACASLAKHFDIFRTVFAEASGELY 2854
 QY 2783 QVWLSCLDPIQVETEDNINATNEFLDEFAKEPVRLGHPLIRETII-KQTKSMRVIMR 2841
 Db 2855 QVWLSCLDPIQVETEDNINATNEFLDEFAKEPVRLGHPLIRETII-KQTKSMRVIMR 2914
 QY 2842 ISHALYDGLSLHVVYKRLHMYNLSRLLPPLPHQFSRYMOYTDAGRESHGFWRDVIONTPM 2901
 Db 2915 LSHALYDGLSLHVVYKRLHMYNLSRLLPPLPHQFSRYMOYTDAGRESHGFWRDVIONTPM 2974
 QY 2902 TILSDTVVDGNDATCK-----ALHLSKIVNIPSOVLGSGNSNITQATVFNAA 2949
 Db 2975 -----TWIKGNNTTPPPPPQOQSTPSGAHASKVVTIPTQA--NTDSRITRATIFTA 3026
 QY 2950 CALVLSRESKSDVGVGRIVSGRGLPVEYODIVGCTNAVPVRA---HIESSDYNQLLH 3006
 Db 3027 CALMLAKEDSDVGVGRIVSGRGLPVEYODIVGCTNAVPVRA---HIESSDYNQLLH 3086
 QY 3007 DIQDQYLLSPHETIGFSDLKRNCTDWPE--AITNFSCTIYHNFEPHESQEQORVEM 3064
 Db 3087 EMQEQYLLSPHETIGFSDLKRNCTDWPE--AITNFSCTIYHNFEPHESQEQORVEM 3146
 QY 3065 GVLTKEVNIEMDEPLDIAIAGEVDPDGLKLVTVIAKTLQFGRKVRVHLEEVSKTEG 3124
 Db 3147 GVLTKEVNIEMDEPLDIAIAGEVDPDGLKLVTVIAKTLQFGRKVRVHLEEVSKTEG 3206
 QY 3125 L 3125
 Db 3207 L 3207

RESULT 3

AAR44929

ID AAR44929 standard; Protein; 15281 AA.

XX AAR44929;

XX AAR44929;

DT 08-JUL-1994 (first entry)

DE T. niveum Cyclosporin synthetase.

XX

Enzyme; cyclosporin; synthetase-like activity; Tolypocladium niveum;
 T. inflatum GAMS; biosynthesis; vector; cyclosporin synthetase.
 Tolypocladium niveum.
 EP578616-A.
 12-JAN-1994.
 05-JUL-1993; 93EP-0810474.
 09-JUL-1992; 92AT-0001403.
 08-MAR-1993; 93AT-0000437.
 29-APR-1993; 93CH-0001310.
 04-MAY-1993; 93CH-0001375.
 (SANO) SANDOZ LTD.
 (SANO) SANDOZ PATENT GMBH.
 (SANO) SANDOZ-ERFINDUNGEN VERW GES MBH.
 Leitner E, Schneider E, Schoergendorfer K, Weber G;
 WPI: 1994-010432/02.
 N-PSDB; AAQ54386.
 Isolated DNA sequence - which codes for enzyme having cyclosporin
 synthetase like activity
 Claim 1; Page 41-84; 93pp; English.
 This sequence represents an enzyme which has cyclosporin synthetase-
 like activity. This sequence was isolated from Tolypocladium niveum
 (formerly known as T. inflatum GAMS). This enzyme catalyses the
 peptide biosynthesis of cyclosporins and structurally related
 molecules. This sequence may be used for the production of
 cyclosporin by transforming a vector containing this sequence in
 to a recombinant host. This allows effective production of anti-
 biotic cyclosporin or its derivatives.
 Sequence 15281 AA;
 Query Match 31.0%; Score 5000; DB 15; Length 115281;
 Best Local Similarity 32.4%; Pred. No. 0;
 Matches 1233; Conservative 440; Mismatches 867; Indels 1264; Gaps 66;
 QY 482 SOPLEQDTLIHHEMLKAVSHSPTKTAQAWDGMWYSBELDNVSRSLAVHKSGLRAQQ 541
 Db 11581 NQPDYPRASVIDVFEQVASTPKSTAVIDASSQLTYTDELSSOLATWLR-QVTYPE 11639
 QY 542 AIPYVFEKSKVVIASMLAVLKSNAFTLIDPNDPPARTAQVVTQT--RATVALTS---- 595
 Db 11640 ELVGLAPRSCETIIAFLGIIKANLAYLPDYNAPAGRIETILSSLPGNRLILLGSDTQA 11699
 QY 596 -KLHRETQKLVGRVVDDELLOSQS-ASDFFSLTSQDLAYVIFSGSGDPAKIMI 653
 Db 11700 VKLHANSV-----REFRISDALVESGSPTELSRPAQSLAYVMTSGSGVPGKGMV 11754
 QY 654 EHRAFSSCALKFGASLGINSDFRALQ-----FGTHAFAGCLLEIMTLLNGG---CV 702
 Db 11755 EHRGIT-----RLVKNNSNVVAKQAAAAHAHLSNTAFDASSWEIYAPLNGGTVCVI 11806
 QY 703 CIPSDDDDRMNSIPSFIRYNVNMWMTSPSYMG---TFSPEDVPLGLATLVGEQMSVVN 759
 Db 11807 DYYTTID-IKALEAVFKQHHRGAMLPALLKQCLVSAPTMISSLEILLPAAGDRSSQ-D 11864
 QY 760 AIWAPKL---QLLNGYQSESSICFASNM-STEPNNMGAVG-----AHSWIDPDNDIN 810
 Db 11865 AILARRVGSVYINAYGPTENTVLTSTIHNGENAEFSGVPTGNVSNVSGAFVMDON--Q 11922
 QY 811 RLVPITGAVGELVIESPGIARDYIVPPPEKSPFFTDIPSYPANTF---PDG--AKLYR 864
 Db 11923 QLVSGAVIGELVVTGDLARGY-----TD--SKLRVDRFIYITLIDGNRVRAYR 11968

QY 865 TGLARY-ASDGSIVCLGRIDSQVIRGQVEIGAIETHL-RQMPDDLIVVEATKRSQ 922
DB 11969 TGDVRHRPKDQGEFFGRDQOQIKIRGHRIEPAEVEQALARPASIDSAAVTLQTDDEE 12028
QY 923 -----SANSTSLIAFLIGSSYFGNRPSSDAHILHDATKA-----INKLEOVL 965
DB 12029 PELVAFSLGNANGTNGV-----NGVSDQEKIDGDEQHALLMENKIRHNLQALL 12078
QY 966 PRHSIPSEYICMLELPRATGKIDRRRLRMGKDILDKOTQGAIVQOAPAPIPVFADTAA 1025
DB 12079 PTYMIPSRITHVDLPVNGKIDRNELAVR-----AQATPRTSSVSTVAP 12125
QY 1026 K--LHSTWOSLGIDPATVNWGAT--FFELGGSNITAIKM-VNMAKSVGMKLVSNYQH 1080
DB 12126 RNDIETIEKEFA-DILSVRGVTDNFFDLGGHSLIATKLAARSLRDLTRVSRVDFDT 12184
QY 1081 PTLAGISAVVKGDPLSTLTPKSTHEGPEOSYSGRLWFLDOLDYVGSILWYLIPYVRMR 1140
DB 12185 PVYQOLAASIQOQSTPHEALPALSHSGPVQOQSFQAQRLWFLDRFNLAAMYIMPFGRRLR 12244
QY 1141 GPNVVDALRRALAEQRHETLRTTFEDQDGVGVQIVHEKLSBEMKVIDLCSGDLDPFEV 1200
DB 12245 GPLRVDALQALRALEERHELLRTTFEEDQGVGMQIVHSPMRDIDCVDISGANED-LAK 12303
QY 1201 LNOEOTTFNLSSEAGHRATLLRLGEDDHLITIVMHIIISDGHSIDLRLDLNQLYSAL 1260
DB 12304 LKEEQAPFNLSTEVAMRVAFKAGENHHILSIVMHIIISDGHSVDIFQOELAQFYSAV 12363
QY 1261 KDSKDPLSALTPLIOYSDPAKWOK--QFIEQEKOLNYWKOLKOLSSPAKIPDTPARPA 1318
DB 12364 R-GHDPLOSQVPLPIHYRDFAVWQODKQAVHESQLOQYIEQLADSTPAEILSDENRPE 12422
QY 1319 LLSGDAGCVHVTIDGELYQSLRAFNEHNTTSFVLLAAPRAAHYRLTAVEDAVIGTPIA 1378
DB 12423 VLSGEAGTVPIEDEVEYKLSLFCRNHQVTSFVLLAAPRAAHYRLTGAEDATIGTPIA 12482
QY 1379 NNRPELEDLIGCFVANTQCRIMIDHDTFTGLINQVKATTAAFENEDIPFERVYSAQ 1438
DB 12483 NNRPELEDLIGFVNTQCMRIALEHDFNLSVVRVRRVSTAASAFENQDVPFERLSALL 12542
QY 1439 PGSRLDLSSTPLAQLIFAVHSQKDLGRFKFQGLSESVVPVSKAYTRFDFMEFLHFOETSLKG 1498
DB 12543 PGSRDASRNPLVQLMFVHVSORNLKQLGLEGEPPTATTRFDVEFLHFOQDKGLAG 12602
QY 1499 SVNFADLPKMETVENVVRVFFELRNGLOSSRTPSVILPLTGVITVEKLDVNLVKHV- 1557
DB 12603 NVVFAADLFEPAATIRSVVEVFHEILRRGLDQPDIAISTMPLVDGLAALNSRNLPAVEDIE 12662
QY 1558 -DYPRESSLADVFQTOVSAYPDSIAYVDDSSCLTYTDLDRQSDILAGWLRRSRMPAETLV 1616
DB 12663 PDFATEASVVDVFQTVAMPNDALAVTDTSKUTYAELOQSDHVAWLSKQKLPAESIV 12722
QY 1617 AVFAPRSCETIVAFVGLKANLAYLPDVRSPSARVQDILSGLSGPTIVLIGHDTAPDOI 1676
DB 12723 VVLAPRSSETIVACIGILKANLAYLPDMSNVPEARQAILSEIPGEKVFLLGAGVPIPDN 12782
QY 1677 EVTNVEFVRDRLDANSNADGFEVIEDHSTKPSATSLAYVLYTSGTGRPKGVMIHVRVI 1736
DB 12783 KTADVRMVFTSIDIVASKTOKSYS----PGTRPSASSLAYVIETSGTGRPKGVMEHVRG 12838
QY 1737 IRVTSGCIPNYPSETRMAHMATIAFDGASYEILYSALLFCRTLVCVVDYMTDLARALKOV 1796
DB 12839 ISLVKONA-SRIQSLRMAHVSALAFDASVMEIETLLNGGTLCFISYFTVLDKALSAA 12897
QY 1797 FFRHVNAAASHVTSSSQDPLVRPRLSRTLMPFFLVVTDSTAPDALDA-QGLYQG---- 1851
DB 12898 FSDHRIN-----ITLLPALLKQCL-----ADAPSVLSSLESILYIGDRL 12937
QY 1852 -----VQCYNGYGPTEGVMSTIYPIDSTESPINGVPTIGRALN-NSGAYVVD 1897
DB 12938 DGADATKVKDLVKGKAYNAYGPTENSVMSTIYIEH-ETFANGVPIGTSLGPKSKAYIMD 12996
QY 1898 PEQOLVIGVMGELVVTGDLARGYSKALKDENRFBVHITVNDQTVKAYRTGDRVRYIGD 1957

DB 12997 QDOQLVPAGVMGELVWAGDGLARGYTDPSLNTGRFIHITDGKQVQAYRTGDRVRYRPRD 13056
QY 1958 GLIEFFGRMDTQFKIRGNRIETSAEIEAALLRDSVRDAAVVLQONEDQAPAILGVVYADH 2017
DB 13057 YQIEFFGRLLDQOQIKIRGHRIEPAEVEQALLSDINDAVVVSQNK-EGLEMVGYITTOA 13115
QY 2018 DHSENDKQGSANQVEGQDHFESGMYSIDIGEDIPSTIGDSFGKWTSMYDGSQIDFDEMHE 2077
DB 13116 AQSV-DKEEASNKVQWEAHEFDSTAYANIGGIDRDALGQDFLSWTSMYDGSILPREMQE 13174
QY 2078 WLGETRRTLHDNRSLGNVLBIGTSGMILNLD--SRLESYVGLPEPSRASAFAFNKATES 2135
DB 13175 WLNDTRSLRLDNOPPGKVLBIGTGMVLENKGVLEQSLQYAGLEPSRSVTAWNKALET 13234
QY 2136 IPSLACKAKVQVGTATDIGOVDDLHPDLVVLNSVIOYFPPSEVYLAETADTLHLPNVQRI 2195
DB 13235 FPLSASARVHVGTAEISSIDGRLSVLVINSVAQYFPPSREYLAELANTLIRLPGVKRI 13294
QY 2196 FFGDVRSQATNEHFLAARAIHTLGKNAKDDVRQKMAELEDMEBELLEVPFAFTSLKDRF 2255
DB 13295 FFGDMRTYATNKDFLVARAVHTLGSNAKAMVRQVAKLEDDBEELLVDPAFTSLSDQF 13354
QY 2256 PGLVEHVEILPKNWEAVNELSAYRYAAVHVGRSL---GDE-----LVLPEKDDWIDFQ 2307
DB 13355 PDEIKHVEILPKRMAATNELSSRYAAVITHVGHQMPNGEDEDKQMAVKDINPKAWYDFA 13414
QY 2308 ANOLNKSGLDLKSSDAA--INAVSKIPPEITAFERQVVASLNSNID-----EWOL 2357
DB 13415 GTRMDQALLQLLQDRQDDVVAVSNIPYSKTIMEHRLSQSLDDDEDGTSAVDGTAN-I 13473
QY 2358 STIRSSAEGDSSLSVPDIFRIAGEAGFRVEVSARQMSONGALDAVPH-----HCCSQG 2411
DB 13474 SRTQSRAKECPALSVADLIEIGKIGIGPEVEASWARQHSQRGGLDPAVPHREPPRH--SG 13530
QY 2412 RTLNVNPTDHLRGSDLLTNRLQRLONRRIALEVRLRLSLPSYMIPIVNLVLDKMLP 2471
DB 13531 HVNFRPTEHKGSSSLTNRLHLLQSRLEAKVRERLQSLPPYMIPIRITLLDQMLP 13590
QY 2472 NANGKVDKRLSRRKVVVPKQOQTAAPLTPFPISEVEVILCEEATEVFGMKVDITDHFENL 2531
DB 13591 TSNKGVDRKLAQARVIPSAASTLDFVAPRTEIEVULCEEFTDLGLGVKVGITDNFEL 13650
QY 2532 GGHSLATKLISRIDORLKVITVKOVDPHPFADLASVTRQG----- 2574
DB 13651 GGHSLATKLISRLRDLAGITVKOVDPQVLADLAASITLOGSSRRHSRPSLPYEGPVE 13710
QY 2575 ----- 2574
DB 13711 QSPAQRLWFLDQFNIDALWYLPFALRMGRPLQVDALAAALVALEERHESLRTTFEERD 13770
QY 2575 ----- 2574
DB 13771 GVGIOVVQPLRTTKDIRIIDVSCMRDDDAVLEPLQEQOQTPDLASEPGHVRALLKLGK 13830
QY 2575 ----- 2574
DB 13831 DHILSVMHIIISDGNWSTEVQLRELQFQYLAAKSGKAPLSQVAPLTIQYRDFAVWQREE 13890
QY 2575 ----- 2574
DB 13891 QVAESORLDYWKQLADSSPAELLADYTRPNVLSGEAGSVSFVINDSVYKSLVFCRSR 13950
QY 2575 ----- 2574
DB 13951 QVTTFTLLAAFPRAHYRWGTGSDATIGTPIANRNPPELENLIGCFVNTOCMRITIGDDE 14010
QY 2575 ----- 2574
DB 14011 TFEESLVQVQVRSSTTATAFENQDVPFERIVSTLSAGSRDTSRNLPLVQLLFAVHSQOGLRIQ 14070
QY 2575 -----LG 2576

Db 14071 LBGVDEPVLSTVSTRFDLEFHAFOEADRLNGSVMFATDLFQPETIQGFVAVVEVLQRG 14130
QY 2577 LQOPVS----- 2582
14131 LQOPSPATMPLAEGIAQLRDAGALOMPKSDYPNALSIVDFQOQAMASPSTVAVTDS 14190
QY 2583 ----- 2582
14191 SKLTYAELDRLSQAASVLRROOLPAETMVAVLAPRSCETIIIAFLAILKANLAYMPLDVN 14250
QY 2583 ----- 2582
14251 TPSARMEALISSVPGRRLLILVSGSVRHADINVPNAKTMLSIVTGTDAIGTPEPLVVRP 14310
QY 2583 ----- 2582
14311 SATSLAYVIFTSGTGKPGVMVEHRAIMRLVKDSNVVTHMPATMAHVTHIAFDVSLF 14370
QY 2583 ----- 2582
14371 EMCATLLNGTLVLCIDYLTLDSTMLRETFEROVRAAIFPPALLRQCLVNNPDAIGMLE 14430
QY 2583 ----- 2582
14431 AUYVAGDRPHSRDARATOMLAGPRVYNAYPTENAILSTIYNIDKHDPVYVPIGSAVS 14490
QY 2583 ----- 2587
14491 NSGAYVMDNRQQLPPGYMGLVVTGEGVARGYTDASLDTDRFVTVTDGQORAYRTGD 14550
QY 2588 ----- 2587
14551 RVRYRPGFQIEFFGRGLDQOAKIRGHRVELGEVHALLSENSVTDAAVVLRTMEBDPOL 14610
QY 2588 ----- 2587
14611 VAFVTTDHEYRSGSSNEEDPYATQAAGDMRRLSLLDYVWVSRVTILRMPNLNANK 14670
QY 2588 -DRS-----AHMAPRTEAILECFKVLGVQVITDNFDFLGGHSL 2629
14671 VDRKDLARRAQMTPTASSGGVHVAPRNETEAICDEFETILGVKGVITDNFDFLGGHSL 14730
QY 2630 MATKLAVRIGHRLDITVSKVDHPVLPOLALDNLV-----OSKTNEIVGGREMAE 2683
14731 LATKLAARLSRMGRISVKOLFDDPVVPSLAKLEQQQGFSGEDESSTVGIV----- 14783
QY 2684 YSPFOLLTFDEPEEFMASIKPOLEL-QEIIODIYPSTOMAKAFILFDHTTARPPFPV 2742
14784 --PFOLLPAEMSERIQDVVQIENGHSTPLDMPATQTOFFLHDKATGHPATPPLFS 14841
QY 2743 IDPSTSEPDAAGLIKACESLVNHLIDIFRTVFAASGELYQVVLSCLDLPQIVIEDNI 2802
14842 LDFPETA--DCRRLASACAALVOHFDIFRTVFSRGGRYQVLAHLDPVEVIEQEEL 14899
QY 2803 NTATNEFFDEFAKEPVRGLGHLIRITIKQT-KSMRVIMRISHALYDGLSLHHVVKLHM 2861
14900 DEVALALHADQOQLRUGRAMLRITAILKRGAKRLVLRMSHLSYDGLSLHHVNALHA 14959
QY 2862 LYNGRSLPLPHPSRYMOTADRGSGHGFVRDVTQNTPTILSDDT-VVDGNDATCKAL 2920
14960 LYSRDLHAQAPFGLYMHMHASRAEGYNFWSILQSSMTSLKRSVGALEAMTPSAGTW 15019
QY 2921 HUSKIVNTPSQVLRSSNIITQATVFNACALVLRSDSKDVFGRIVSGROGLPVEVQ 2980
15020 QTSKSIRIPPAALK--NGITQATLFTAAVSLLLAKHTKSTDVFGFVSVGRQDLSINQ 15076
QY 2981 DIVGCTNAPVRAHI-ESSDYNQLLHDIDQOYLLSLPHETICGFSDLKRCNTDWEA 3039
15077 DIVGPCINEVVRVIDEGDGMGLLRAIQOYTSFRHETILGLQEVKENCNTDWDATKE 15136
QY 3040 FSCCITYHNFYHPEFSQFQOQVEM----- 3064
15137 FSCCITAFQNLNLHPEAIEGQIRLEGLPAKDQARQANGHAPNGTNGTNGTNGANCT 15196

QY 3065 -----GVLTKFVNIEM-----DEPLYLAIAGEVEPDGAGLKVTVIK 3102
Db 15197 NCTNGTNGTHANGSNGVGRSDSNVVSAGDQAPVHDDIVGPEPDGS-VKIGIGAS 15255
QY 3103 TOLFGRKRVRHLLLEVKSTFEGLN 3126
Db 15256 ROILGEKVVGSMNLCELTMLALS 15279
RESULT 4
AAB07581
ID AAB07581 standard; Protein; 2841 AA.
XX
AC AAB07581;
XX 20-OCT-2000 (first entry)
XX
DE Protein encoded by the bleomycin (BLM) gene cluster ORF31.
XX
KW BLM gene cluster; bleomycin gene cluster; polyketide metabolite;
KW bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;
KW thiazoline; bithiazoline; microbial metabolite; sugar.
XX
OS Streptomyces verticillus.
XX
PN WO2000040704-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00445.
XX
PR 06-JAN-1999; 99US-0115435.
PR 05-FEB-1999; 99US-0118848.
PR 05-JAN-2000; 2000US-0477962.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Shen B, Du L, Sanchez C, Chen M, Edwards DJ;
XX
XX WPI: 2000-465974/40.
DR N-PSDB; AAS58472.
DR
XX
PT New bleomycin gene cluster components useful for peptide and/or
PT polyketide metabolites, especially bleomycin, production and for
PT chemically modifying biological molecules -
XX
XX Disclosure; Page 138-153; 162pp; English.
XX
CC AAB07580-89 represent proteins encoded by open reading frames (ORFs)
CC 8 to 30 of the BLM (bleomycin) gene cluster. The proteins encoded
CC by the gene cluster are useful for producing peptides and/or polyketide
CC metabolites, especially bleomycin or bleomycin analogues. They are
CC also useful for chemically modifying biological molecules to produce
CC branched methyl groups, and for coupling amino acids and fatty
CC acids. They may be reacted with an apo-carrier protein and coenzyme A
CC to produce a holo-carrier protein. The BLM gene cluster or catalytic
CC domains can be used individually or collectively to produce
CC thiazolidine, thiazoline, bithiazoline and bithiazoline-containing
CC microbial metabolites. The BLM gene cluster may also be used to produce
CC sugars.
XX
SQ Sequence 2841 AA;
Query Match 13.3%; Score 2146.5; DB 21; Length 2841;
Best Local Similarity 28.9%; Pred. No. 3 9e-163;
Matches 722; Conservative 394; Mismatches 1025; Indels 357; Gaps 91;
QY 235 FWOTHLNDLNASVFFHLSDLHVMVNPNTTAA--HRTTFPLSOKALS-----NSAIC 283
Db 539 YNKRAL-DCAPSVLRPLRPMDH---PRPAVQSERGETVGFALPDALVAALKLREGQCATLF 594
QY 284 RT---ALSTLLSRYSDEALFGAVTEQSLPDKHYLADGTYQTVAPLRVHCQSNLRASD 340

QY	739	EDVPGIATL-----VIVGEQMSVSNAINAPKIQLLNGYQSSSSSICFASNMSSTEP	790
Db	713	KDAGMFOTLRLHLIIGGDALPHVIVSKVQA--SPSLSLWNGYGTENTTSTFSLDREY	770
QY	791	NN---MGRVAG-AHSWVIDPNDINRLVPIGAVGELVIESPGIARDYIVRPPPKSPFFTD	846
Db	771	GGSTIPCKPIGNSTAYIMDEQOC--LOPIGAPELCVGGVGARGVYNLPELTEKOFLED	828
QY	847	IPSWYPANTFPDGAKLYRTGDLARYASDGSIVCLGRIDSQVKIRGORVELGAIETHLRQO	906
Db	829	-----PRPGERIYRTGDLARWLPGDNIEFLGRIDNQVKRGFERIELGETETKL--N	878
QY	907	MPDOLT--TVREATKRSOSANSTSLAFILIGSSYFCGNRPSDAHILHDHDKATKAINKLEQ	963
Db	879	MAEHVTAHVII---RKNKADENECIA-----YF--TADREV---AVSELKRTLSQ	921
QY	964	VLPRHSIPSYICMLELPRATATKIDRRRLRIMGKIDLKQTOGAIQVQAPAPIPVFA--	1021
Db	922	SLPDYMPFAHLIQMDSLPLTPNGKINKKELP-----APQSEAVOPEVAAP	966
QY	1022	--DTAAKLHISWQSLGIDPATVNVGAT--FFELGGNSITAIKM-VNMAHSVGMCDLKVS	1076
Db	967	KTSEKKLABIEGILG-----VKAGVTDNFFMIGGHSKAMMTAKIOEHFKEVPKIV	1021
QY	1077	IYQHTPLAGISAVVK-----GDPLSYTLIPKSTHEGPVEQSYSQRLWFLDQLDVGSILWL	1132
Db	1022	LFEKPTIQELALYLEENESKEEQTEFPIROASYYQHYVPVSPAQRMYILNQLQOANTS	1081
QY	1133	IPAVMRGPNVDALRRALAALQORHETLRTTFEDQDGVGVQIVHEKLESEMKVIDLQ	1192
Db	1082	VPVALLLEGVEVDKRLNENAIQQILNHRHEILRTSFDIMDGEVQTVHKNISFHLAAK--G	1139
QY	1193	SDLDPPFVLNQEOTTPFNLSSEAGWRATLLRGEDDHLITVNMHIIISDCWSIDVLRRDL	1252
Db	1140	REDAEEII-KAFQPFELNRAPLVRSKLVQLEKEKRLHLLIDMHHIITDGSSTGILIGDL	1198
QY	1253	NOLYSAAKDKDPLSALTPLOYSDFAKWKQDQFIEQKQLNYWKQKQKLDSSP-AKIP	1311
Db	1199	AKIYQGA--DLELP-----QIHYKDYAVWHKEQ-TNYQKDEYWLDFVKGLPILDL	1248
QY	1312	TDFAHPALLSGDAGCVHVTIDGELYSLRAFONHNTTSFVVLAAAFRAAHYRLTAVEDA	1371
Db	1249	ADFERPAERSFAGERVMFGDKQITQAIKSLMAETDITMTMFLAAAFNVLLSYASQDDI	1308
QY	1372	VIGTPIANRNRPELDIIGCFVNTQCMRINIDHDTFGTLINQVAKATTAAAFENEDIPFE	1431
Db	1309	IVGSPTAGRTHPDLOGVPGFMVNTGALTAPAGDKTFAQFLEEVKVTASLAQAREHQSYPL	1368
QY	1432	RVYSALOPGSRDLSSTPLAQLIIFAVHSQ-----DLGRFKFOGLSVPPVPSKYATRED	1484
Db	1369	ELIEKL-PLTRDTRSPFLSVNMNMNMTIPSURLGDLKITSYSMLHHV-----AKFD	1420
QY	1485	MEFHLFQETDSLKGSVNFADLEFKMETVENVVRVFFELRNGLQSSRTPVSIPLPTDGV	1544
Db	1421	LSLEAVEREEDIGLSPDYATALKFDETIIRWSRHFVNIK--AAANPNVRLSDVDLLS	1477
QY	1545	TLKLDVLNVKHVDYPRESSLADVOTQVSAYPDSLAVVDSSCRLTYETLDRQSDILAG	1604
Db	1478	SAETAALLEERHMTQITEATFAALLEKQAQOQTPDHSAGVAGGNLLTYRELDEQANLAHH	1537
QY	1605	LRRSMPAETVLVAFAPRSCETIVAFGVVLKANLAYLPLDVRSPSARVODILSGLSGPTI	1664
Db	1538	LRAQAGNEDIVAIVMHSRSEVMVSMVMKAGCAPFLPDPTPGERIYSLBH-SGAKF	1596
QY	1665	VLIGHTAPDPIETVNVFVRIRDALNDSNADG---FEVIEHDSKPSATSLAYLYTSG	1721
Db	1597	AVVNE-----RNMTAQOGEIVTSLDDGKWRNESKERPSSISGSRNLAYVIYVTS	1647
QY	1722	STGRPKGVMIHRVILRTVTSGCIPNYP-----ETRMHAMATIAFDGAS	1766
Db	1648	TTGPKGVQIHR-----NLTNVFSVSEAGLTRRADGNDKTVLLSSYAFDLGY	1698

Db 1162 LRTFRLEGGRSYQVPAVSISIEREQFGEEGLI-----ERIQAIIVQPEDE 1210
QY 1213 SEAGWRATLLRLGEDDHILITVMHHIITSDGNSIDVLRDLNQLYSAAKDSKPLSALTP 1272
Db 1211 RGPLRLVNLQLAEDDHVLVLQVHHIVSDGMSQVMVEELVQLV--AAYSQGLDVV--LPA 1267
QY 1273 LPIQYSDFAKWKQD--QFIEQEKOLNWKKOLKQSSPA-KIPTDFARPALLSGDAGCVHV 1329
Db 1268 LPIQIADYALMQRSMWAGERKRLAYWTGLLGEQVLELFPDRPRPAROSHGAQLGF 1327
QY 1330 TIDGELYQSRAFACNEHTTTSFVYLLAAAFRAHRLTAVEDAVICTPIANRNRELEDDII 1389
Db 1328 ELSRELVEAVRALAQREGASCFMILLASFOALLRYSSQADIRVGVPIANERNVETERLI 1387
QY 1390 GCFVNTQCMRINIDHDTFTGLINQVATTTAAENEDIPERVVVSALQPGSRDLSSTPL 1449
Db 1388 GFFVNTQVLKADLDGRMGFDLELAQAQRALEAQAHQDLPEQLEVALQP--ERNASHNPL 1446
QY 1450 AOLIF-----AVHSOKDLGRFQOGLESVPVPSKAY----TREDMEHFLFOETDSIKG 1498
Db 1447 FOVLFNHQSEIRSVTPVQVLEDLRLGL-----AWDQTAQFDLTLDIQEDENGIIWA 1498
QY 1499 SVNFADFLKMETVENVVRVFFETLRNGLOSSRTPVSILPLTDGIVTLEKLDVLNVKHDV 1558
Db 1499 SFDYATDLEDASTVERLAGHWRNLRGIVANPRQRLGELPLDAPERRQTLSEWNPQRE 1558
QY 1559 YPRESSLADYFOTQVAPPSLAVVSDSCRLTYTELDRQSDILAGWLRRRMPAETLVAV 1618
Db 1559 CAVOGTTLQORFEQARQPOQVAILIDQRLSYGELNARANRLAHCLIAICVGVADVPVL 1618
QY 1619 FAPRSCETIVAFVGLKANLAYLPDVRSPSARVODILSLGSGPTIVLI--GH--DTAPPD 1675
Db 1619 ALERSLDWLVGLLAILKAGAYLPDPAAPERLAHLDD--SGVRLLLTOGHLLERLP-- 1675
QY 1676 IEVTNVEFVRIDALNSDADGFEVIEHDKTP-----SATSLAYLVLTSGTRPKGVMI 1731
Db 1676 -ROAGVEVLAI-----DGLVLDGYAESDPLPTLSADNLAVYVTSGTGPKGTLL 1725
QY 1732 EHRVIR--VTSICIPNYPSETRMAHWATIDGASYEIIYALLFRTLCVVDYMTILD 1789
Db 1726 THRNALRLFSATENWF--GFDERDVTWTLFHSATDFSVMEIFGALLIGRVLIVPQWVSRS 1784
QY 1790 ARALKOVFFREHVNAASHVTS-----SQDPLRVPRRLSRITLMFFELVVVTDSTAP 1840
Db 1785 PEDFYRLCREGTVTLNTPSAFKQLMAVACSADMATQQP--ALRYVIF-----GG 1833
QY 1841 DALDAQGLYQVQCY-----NGYPTENGVMSTIYPIDSTESFNG---VPIGRALN 1889
Db 1834 EALDQSLRPWFQRFGRQFQPLQNMVIGITETTVHVTYRPV--SEADLKGGVLVSPIGGTIP 1891
QY 1890 NSGAYVVDPEQQLVGIGVMGELVVTGDGLARGYSDK--ALDENRFV--HITVNDQTVKAYRT 1947
Db 1892 DLSWYLLDRNLNPVRGAVGELYIGRAGLARGYLRRPCLSATRFVNPFPFGGAGERLYRT 1951
QY 1948 GDRVYRIGDGLIEFFGRMDTQFKIRGNRIESAEIEAALLRDSVRDAAYVVLQONEDQAP 2007
Db 1952 GDLARFQ--ADGNTIEYIGRIDHQVKGVRGRIELGEIEAALAGLAGVRD--AVVLAHDGVGGT 2009
QY 2008 EILGFVVADHDHSENDKQSANQVEGQDHFEQSMYSDI-----G 2047
Db 2010 OLVGYYVAD-----SADAEARLESRLSKRHLPDYVMPVPAHMLLERMPLTVNG 2059
QY 2048 EIDPSTI-----GSDFKG-----WTSMYDSQIDFDEMHEWLGTEYTLH 2087
Db 2060 KLRQALPQDASLSQAYRAPSDEQRTAAIWAELIGVERVGLDD----- 2106
QY 2088 DNKSLGNVLBIGHTGSGMIFNLDSRLSYVGLFSPRSAAAFVNKATESIPSLAGKARQV 2147
Db 2107 -----NFFELG--GHSLLLLMLKERIGD-----TCQATLISISQLMTHASVAE 2146
QY 2148 GTATDQVDDHPLDVLVNS-----VIOYFSPSEYLAETDLIHLPNQWRIFFGDVR 2201
Db 2147 QAACIEGQARE--SLIVPLNGRREGSPLFMFHPS-----FGSVH 2183

QY 2202 SQAT-----NEH-----FLAARAIHTLGNATK--DDVRQKMAELEDMEELLVPEAFFTSL 2251
Db 2184 CYKTLAMALRDRHPVKGVVCRALLGAGREVPEWDDMVAEYAE----- 2225
QY 2252 KDRFPGLVHEVEIILPKMNEAVNELSAYRYAAVHVHRSGLGDELVLVPEKD----- 2301
Db 2226 -----OLLOHEPGEVFNLAGW-----SLGNLAMDVAARLEORGRQVAF 2264
QY 2302 -DWIDFOA 2308
Db 2265 VGWIDAPA 2272

RESULT 7

AAU33611
ID AAU33611 standard; Protein; 4342 AA.

XX AAU33611;

XX 14-FEB-2002 (first entry)

XX Pseudomonas aeruginosa cellular proliferation protein #55.

XX Antisense; prokaryotic cellular proliferation protein;

XX antibiotic; antibacterial; drug design.

XX Pseudomonas aeruginosa.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 22-NOV-2000; 2000US-253625P.

PR 27-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

(ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

DR N-PSDB; AAS51470.

XX New polynucleotides for the identification and development of
antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 5107; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
prokaryotic cellular proliferation, their use in identifying the
genes, their use in the discovery of novel antibiotics, the essential
genes themselves and the encoded proteins. The prokaryotes used are
Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
invention is also useful for the identification of potential new targets
for antibiotic development. The antisense nucleic acids can also be used
to identify proteins used in proliferation, to express these proteins,
and to obtain antibodies capable of binding to the expressed proteins.
The proteins can be used to screen compounds in rational drug discovery
programmes. The antisense nucleic acid sequence is also useful to screen
for homologous nucleic acids which are required for cell proliferation in
a wide variety of organisms. The present sequence represents an
essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
of the printed specification, but was obtained in electronic

Db 2554 RFPDPFAEGRLRYTGDVRL-CDNGQVEYVGRIDHQVKIRGFRLEGEIEARLLEHP 2612
Qy 1991 SVRDAAVYVQONEDQAEIILGFV---VADHDHSNDKQSQANQVEGQWDFESGMYSDIG 2047
Db 2613 QVRE-ALVLAIDSPSGKOLAGYASVAEQD---EDAQA-----2648
Qy 2048 EIDPSTIGSDFKGMTWYDGSQIDFDEMHEWLGTTRTLHDNRSLGNVLEIGTSGMILF 2107
Db 2649 -----LREALKT-----2655
Qy 2108 NLSRLSEYVGLPERSAAAFVNKATESIPSLA-GKAKVQVGTATDIGQVDDLHPDLVLV 2166
Db 2656 HLKQOLPDY--WPAHLL-----LLASPLTANGKLDRLRALPAD-----PAL 2696
Qy 2167 NSVIQYFSPSEYLAETIHLFPNVQRIFFGD-----VRSQATNEHF--2209
Db 2697 NRQAYEAPRSVLEQOLAGVWREVLNVERVGLGDNFFELGDSILSIQVSRARGLIHFS 2756
Qy 2210 -----LAARAIHLGRNATKDDVR-----QKMAELEDMEELLY 2243
Db 2757 PRDLFQHTVQSLAAVAARHSQASQAEQGVQGSALTPIQHWFDFDLPLARREHWNQALL 2816
Qy 2244 EPAFFTSLKDRFPGVVEHVEILPKNME-AVNELSAYRYAA-----VVHVRGSLGDELV 2295
Db 2817 QPROAIDL-----GLLRKSLQRLVEQHDALRLAFQVDGGEWLAQHRPLREDELL 2865
Qy 2296 --LPVEK-DWMIDFOANOLNKSLSL---GDLKSSDAAIMAVSKIPPEITAFERQVVASL 2348
Db 2866 WHVPVQSFDECAELFAKA--QRSLEDEQGPLLRA-----VLVDGPAGEORULLAI 2913
Qy 2349 NSNIDE---WOL-----STIRSAEGDSSLVDPDIFRIAGEAGFRVEVSARQWSONGA 2399
Db 2914 HHLVVDGVSRRVLLEDLQOYIROFAEG-----AEPALPAKTSAFRDWA--GR 2958
Qy 2400 LDVAFHCCSO-----GRVLNFPETHHLRGSD-----LLTNRLPQ 2435
Db 2959 LQA---YAGESLREELGWQARLGQGVPEWPCDRP-QGDNREALAESVLSRLDPQTRQ 3014
Qy 2436 RLQNRRIA--TEVRERLSRLPSYMI-----PSNIVILD-----2467
Db 3015 LLOQAPAAAYRTQVNDLLLTALARVLCRWSGQPSTLVQLEGHREALFDDIDLTRSVGWET 3074
Qy 2468 -----KMPLNANGKVDKRLSRKVVPK-----QOTAAPLPTPISE 2505
Db 3075 SAYPLRUTPAQSPGE-SIKAIKESOLRAVPHKGLGYVRLYADPAVRQANMAALPTAPIT- 3132
Qy 2506 VEVILCEATEVFQMKVDITDHFENLGGHSLLATKLISRIDQRLKVRITVKDVFDPHVFVA 2565
Db 3133 -----FNYLGO-----FDQS-FA 3144
Qy 2566 DLASVIRQGLQOPVSDGQODRSAMAPTETETAILCDEFKVLGVQVQITDNFFDLG 2625
Db 3145 D-----ALFQLDIOPGT-----PIHDEQAPLPNELS--VDGQV-----YG 3177
Qy 2626 GHSMLMATKLAVRIGHRDLTIVSKVDVDPHVPFLQALTA-----LDNLVQSTNEIVGREGM 2681
Db 3178 GELV-----LRWTYS--RERYDARTVNELAQYLAELQALIEHCELDGAGGL-- 3222
Qy 2682 AEXSPFOLLTFTEPEEFMASEI-KPOLELQEI-----IODIYPTOMOKAFLEFHTHTARPR 2736
Db 3223 -----TPSDFPLAQLSQAQLDALAVPAGEIEDYVPLTPWQEGLLL-HTLLEPG 3269
Qy 2737 PFV-----PFYIDFPSTSEPAAGLIKACESLVNHLIDIFRTVFAESAGE-LYQVVLSCLD 2790
Db 3270 TGIYVMDQRYRIDSPLDPERFAA---AWQAVVARHEALRASFWNAGETMLQVHKPGR 3325
Qy 2791 LPTQVETEDININTATNEFIDEFAKEPVRGLHPLIRFTTIKQKSRMVRIMRSHALYDGL 2850
Db 3326 TRIEFDLWSLEPDGHEERLQALHKREAG-----FDLEEQPFHLRLRLGEARYWFM 3380
Qy 2851 SLEHVVRKHLMLYNG--RSLI-----PPHQFSRYMQYTA---D 2883

RESULT 8

AAO22159

ID AAO22159 standard; Protein; 4999 AA.

XX AAO22159;

XX AC AC

XX DT 03-OCT-2002 (first entry)

XX DE Ramoplanin biosynthetic ORF 14 protein.

XX KW Ramoplanin; ramoplanin biosynthetic pathway polypeptide; antibiotic;

XX KW biosynthesis gene cluster; bioengineering; peptide synthetase module;

XX KW adenylation domain; hydroxyphenylglycine; HPG; antibiotic precursor;

XX KW chlorinate; lipopeptide.

XX OS Actinoplanes sp.

XX PN WO200231155-A2.

XX PD 18-APR-2002.

XX PF 15-OCT-2001; 2001WO-CA01462.

XX PR 13-OCT-2000; 2000US-239924P.

XX PR 12-APR-2001; 2001US-283296P.

XX PR 24-JUL-2001; 2001US-0910813.

XX PA (ECOP-) ECOPIA BIOSCIENCES INC.

XX PI Farnet CM, Zazopoulos E, Staffa A;

XX PI WPI; 2002-435445/46.

XX DR N-PSDB; AAL40781.

XX PT Novel isolated ramoplanin biosynthetic pathway polypeptide useful for

XX PT chemically modifying biological molecule that is a substrate for a

XX PT polypeptide encoded by a ramoplanin biosynthesis gene cluster -

XX PS Claim 14; Page 169-186; 212pp; English.

XX CC The invention relates to an isolated ramoplanin biosynthetic pathway

XX CC polypeptide selected from a polypeptide of open reading frames (ORF)

XX CC 1-32. The isolated polypeptides are useful for chemically modifying a

XX CC biological molecule that is a substrate for a polypeptide encoded by a

XX CC ramoplanin biosynthesis gene cluster, by contacting the biological

XX CC molecule with the isolated polypeptide, where the polypeptide chemically

XX CC modifies the biological molecule. The method comprises contacting the

XX CC biological molecule with at least two different polypeptides encoded by

XX CC ramoplanin ORFs 1-31. The polypeptides are useful for directing the

XX CC biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated

QY	1680	NVEFVRIRDALNSNADGFEVIE-----HDSKTPSATSLAYLYTSGTGRPKGVMIHR	1733
DB	2703	GVPCLAII-----DPAIAAEPAPEDGDPDGAAGRPDPDPAYIIYTSGTCTPKGVVSHR	2757
QY	1735	VIIRTVT--SGCIPNYPSETRMAHMATTAFDGCASVEIYSALLFGRTLVCVDMYMTLORAL	1793
DB	2758	NVALLATATRLFCFAGDEWNSVHSAFDFSVWELMGALTHGTRVVVVVYAVSRSPDF	2817
QY	1794	KDVEFRFHNAAASHVTSSQODVLRV--PRLRS:LTLMFFFLVVDSTADPADLDAQGLYQ	1850
DB	2818	WELVVREGVTVLSOTPSAFQALMAAGDDDRDALRFVVF-----GGEALD-PGRLA	2867
QY	1851	G-----VQCYNGYGTENGVMSTIYPIDSTESFINGVPIGRALNNSGAYVVDPEQOL	1902
DB	2868	GWLARRDPKRLVNMGITETVHTTYOHIAPGTT--GSVIGRGLPGFLGYLVLDEALRP	2924
QY	1903	VGIGVMGLVVTGDLGARYSDK--ALDENRFVHITYNDQVTKAYRTGDRVRYRIGDGLIE	1961
DB	2925	VPAGVGEVARGPQVARGYIGRPLGTAERFV--ASFPAPGERMYRTGDVARM--TADGRLV	2982
QY	1962	FFGRMDTQFKIRGNRIESAETEARALLRDSVRDAAVVLQONEDQAPELLGFVVADHDHSE	2021
DB	2983	FAGRSDDOIKIRPRIEPGEVEAVLAAGPGVQAQAVIVRE-----	3022
QY	2022	NDKGOSANQVEGMQDHFESGMYSDIGETIDPSTIGSDFKGWTSMDYGSQIDFDEMHEWLGE	2081
DB	3023	-----DVPG-----	3026
QY	2082	TTRTLHNRSIGNVLEITGTGSGMILFNLDRLSEYVGLEPSRSAAPAVNKATSIPLAG	2141
DB	3027	-----DKRLVAVY-----	3034
QY	2142	KAKVQVGTATDIGQVDDLHPDLVVLNSVIOYFFSSEYLAETLIHLPNVQRIFFGDVR	2201
DB	3035	-----VG-----GDA--	3039
QY	2202	SOATNEHFLAARAITHLKGKATKDDVRQKMAELEDMEELLVEPAFTSLKDRFGLVEH	2261
DB	3040	-----	3039
QY	2262	VEILPKWEAVNELSARYAAYVHVHVRGSLGDELVPVEKDDWDIFQANQLNKSGLDLK	2321
DB	3040	-----	3039
QY	2322	SSDAAIMAVSKIPEITAFERQVVASLNSNIDEWQLSTIRSSAEGDSSLSPVDIFRIAGE	2381
DB	3040	-----ETLRSHAQ-----	3047
QY	2382	AGFRVEVSSARQWSQNGALDAVFHCCSQGRTLVNFTPDHHLRGSDLLTNRP:LRQLQNR	2441
DB	3048	-----QR-----	3049
QY	2442	IATEVRELRSLLPYSNIVVLDKMPLNANGKVDKRLSSRAKVVVPKQQTAABLPTF	2501
DB	3050	-----LPGYLVPSAFVDELRLPTVNGKLDRR-----ALPVPDY	3083
QY	2502	PISEVEVILCEEATEVFGMKVDITDHFNLGGHSLATKLI:SRIDORLKVRI:TKDVFEDH	2561
DB	3084	-----	3083
QY	2562	PVFADLASVIRQGLGLQPPVSDGGODRSAHMAPRTETEA:ILCDEFKVLGFO--VGITDN	2620
DB	3084	-----GRDAGGGRAPANAREEVLACRAEVLGVERVGVEDD	3119
QY	2621	FFDLGGHSLMATKLAVRIGHRLDITTVSKVDVDPHVPVLFQALADNLNLSKNTNEIVGGRE	2680
DB	3120	FFALGGHSLWLSVERL--RROGISVPVRAFTTPTPAGLA-----EAVG--D	3164
QY	2681	MAEYSPQLLFTEDPEEFMAEIKPOL-----ELOEI-----IODIYPSQ	2721
DB	3165	GAUVVPPNLI-----PED--AAELTPEMLPADUTADELAVVAVSVPGGAANIADTVPLAP	3218
QY	2722	MQKAF:LDHTTA-RPRPFV---PFYIDFPSTSEPDAGLIIKACESLVNHLIDFRT-VFAE	2776

Db 2664 -----DGSRTILSLPLDENDEPENDETAVTAENLAYMIYTSCTTQOPGVVVEHH 2715
Qy 1735 VIIRTVSGTIPNY-----PSETRMAHMAITAFDGSAYEYISALLFGRTLCVVDYMTFLD 1789
Db 2716 ALVNL-----CFWHDAFMTAEDRSKAYAGFDASIWEMFPTWSIGAEHLVIEEARLD 2771
Qy 1790 ARALKDVF-----FREHYNASHVTSOSSQVDPVPLRVPRLSLRMLMFFFL 1832
Db 2772 IVRLNDYFETNGVTITPLTQALAEQFMELENTSLRVLLTGGD-----KLKRAVKPYT 2824
Qy 1833 VVTDSTAPDALDAOGLVQGVQCYNGYGTENGVMSTIPIPDSTESFNGVPIGRALNNSG 1892
Db 2825 LV-----NNYGTENTVVATSAETHPEEGSL-----IGRANTANR 2860
Qy 1893 AYVVDPQQLVGVGMGELVVTGDLARGYSDKALDE--NRFRHITVNDQTV--RAYRT 1947
Db 2861 VYILGEGNQVPEGVAGELCVAGRLARGYLNR--DEAKRF-----VADPFVGERMYRT 2915
Qy 1948 GDRVRYRIGDGLIEFFGDMTQPKIRGNRIESAEIEAALLRDSVRDAAVVLOONEDQAP 2007
Db 2916 GDLVKWT--GGGIEYIGRIDQYKVRGYRIELSEIVOLAQLSEVQDAAVTAVDKGNT 2973
Qy 2008 EILGEVVADHDHENDKQGANQVEGWQDFESGMYSDIGEIDPSTIGS-----DF 2058
Db 2974 AIAAYTPE--SADIEALKSALKETLPDIWIFAFWTLNLPVTANGKVDKRALNEPDI 3030
Qy 2059 KGWTSYDQSGIDFDEMHEWLGETTRTLHDNRSLGNVLEIGTSGMIFENLDSRLSEY-- 2116
Db 3031 EAGSGEYKAPTTDMEELLAGIWDVLCMSVGVTDNFFSLG--GDSIKGIQMASRLNQHW 3089
Qy 2117 -----VGLPSSRAAFVYNKATESIPSLAGKAKVQVGTATDIGQVDDLHPDLVLSVI 2170
Db 3090 KLEMKDLNQHT-----IEELQYVERAEGK-----QADQGPVE-----GEVILTIQ 3132
Qy 2171 QYPPSSSEYLAE--IADTLHLPL--NVQRIFFGDRVSQATNEHFLAARAIHTLGNATK 2224
Db 3133 RWFEEKNFTNKHWNQSVMLHAKKGPDPERV---EXTLQALIEHHDALRMVYREG---Q 3185
Qy 2225 DDVQR-----KMAELE-----DMEELLVPEAFPTSL 2251
Db 3186 EDVIOYNRGLEAASAOLEVIQEGQAADYEDRIEREAEALQSSIDLQEGGLLQAGLFOAE 3245
Qy 2252 K-DRFPCLVHEVILPKMMEAVNELSARYAAVVHVRSGLGDELVLPL----- 2297
Db 3246 DGOHLLLAHLVY-----DGVSWRILLEDFAVYITQLEQGNFPLVPKQTHSPAFAERL 3300
Qy 2298 -----VEKDDWIDFOANLNQKSLGDL-----KSSDAAIMAVSKIPFEITAFERQ 2343
Db 3301 QDFANSKAFLEKEYW-----SOLEQAVAAKLPKDRSGDQRMKHTKTIEFSLTAETE 3355
Qy 2344 VVASLNSNIDE 2354
Db 3356 ---QLTTKVHE 3363

RESULT 10.

AAB83971
ID AAB83971 standard; Protein; 1537 AA.

XX AC AAB83971;

XX DT 06-AUG-2001 (first entry)

XX DE Amino acid sequence of a type I polyketide synthase.

XX KW Metabolic pathway operon; polyketide; polyketide antibiotic;

XX KW type I polyketide synthase.

XX OS Unidentified.

XX PN WO20010497-A2.

XX PN 07-JUN-2001.

XX PF 27-NOV-2000; 2000WO-FR03311.
XX PR 29-NOV-1999; 99FR-0015032.
XX PR 07-JUN-2000; 2000US-0209800.
XX PA (AVET) AVENTIS PHARMA SA.
XX PI Jeannin P, Pernodet J, Guerinneau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX DR WPI; 2001-374849/39.
XX DR N-PSDB; AAF90034.
XX PT Collection of nucleic acids from environmental samples, useful for
XX PT identifying e.g. genes encoding polyketide synthases and derived
XX PT antibiotics
XX PS Claim 78; Page 313-318; 356pp; French.
XX CC The specification describes a method for the preparation of a collection
XX CC of nucleic acids from organisms in a soil sample. The method comprises
XX CC milling a dried sample to produce microparticles; suspending these in
XX CC liquid buffer; extraction of nucleic acids from the microparticle;
XX CC passing nucleic acid-containing solution through a molecular sieve;
XX CC passing nucleic acid-enriched fractions through an anion exchange
XX CC chromatography material; and recovering fractions containing purified
XX CC nucleic acids. The nucleic acids are sources for sequences that encode
XX CC either operons involved in a metabolic pathway (specifically polyketide
XX CC synthases) or polypeptides, particularly for production of therapeutic
XX CC or agricultural compounds, especially polyketide antibiotics. AAB83971-76
XX CC represent type I polyketide synthases encoded by cosmid a26gl.
XX SQ Sequence 1537 AA;

Query Match 9.7%; Score 1572; DB 22; Length 1537;

Best Local Similarity 29.3%; Pred. No. 4.8e-117;

Matches 476; Conservative 268; Mismatches 670; Indels 208; Gaps 51;

Qy 1040 ATNVGATPELGGNSTAIKMYNMARSVGMGLKV-----SNIOHPT-----LAG 1085
Db 6 SAVDLOQPLVRMGDLSLMAVOLRN---RIDTDLRVLLPMVRFDLGFSVABLRLDLSLG 62Qy 1086 IS--AVVKGDPISYLLIPKSTHEGPEQSYSGRLMFLDQDVGSLWYLPYAVRMGPV 1143
Db 63 LSERTTVAPEPAAQASVPALSY--PL--SAGQALWFIYRSAPESPAYNIATARAGAF 118Qy 1144 NVDALRRALAAEQRIHETLRTTFEDQDQGVQVIVHEKLSKEMKVIDLGGSDLPFVNLQ 1203
Db 119 DQALRRSLQDLVDRHPALRTTIAESGGAPVQTVHSSVPVDFEIP--CSPD--DEAVLIDG 176Qy 1204 EQTTPENLSSEAGWRATLLRLGEDDHTITVMHHIISDGWSIDVLRDLNQLSAALKDS 1263
Db 177 VFHAPNL--GENCFRSLLVQSGKDQVLAIVVHILADFWSLLVWDELRSIYLA--RTA 233Qy 1264 KPPLSALTPLPIQYSDFAKWKQDF--IEQEKLYNWKKLDKSSPA--KIPTDFARPALL 1320
Db 234 GGP--VPAPPVASFAAFVRWQNELLAGTEGERLWNYWSSQLSGOLPVLNLPDRSPVPQ 291Qy 1321 SGDAGCVHVTIDGELYQSLRAFNCNEHTTTSFVLLAAFRAAHYRLTAVEAVICTPIANR 1380
Db 292 SFRGNSSHFRIEPALTAKLARQNATLHATMAAFQVLLLRWTSQEEILYGLTNGR 351Qy 1381 NRPELEIIGCFVNTQCMRINIDHDTFTGLINQVRATTTTAAFENEDIPERVVSALQPG 1440
Db 352 TQPEFADLVGYFVNPVILRGELSCDPDFNTVARIQTLLGAIEHQEYPARIVERLGP 411Qy 1441 SRLSSTPLAQLFAVHSQKDL-----GRFKFG--LESVPVPSKAYTRPDM 1486
Db 412 LR-----VLFVLQOQPHRIPESVPFMLGSGGRMANGSLTLESAMPRL--QSRFDLD 461

Qy 1487 FHLFQETDLSLGSVNFADFLFKMETVENVVRVFEILRNLQSSRPVSVILPLTDGIVTL 1546

Db 462 LMVETDGGLSAFQYNTDIFDAATIERLSHFVAVLLEGIAENPACPVVDLPL---LSTR 518
 Qy 1547 EKLDVL---NVKHVDYIPRESSLADVFOQYSAVPSDLAVVDSSCRLLTYTDLDRSDTLAG 1603
 Db 519 ERITQLEENWATAAEFSPQ-CVHELFQAQVELIPDAISFGEONLIYRELNGSANRIAH 577
 Qy 1604 WLRSSMPAETLVAVFAPRSCETIVAFGLVKANLAYLPDLVRSARVQDILSGLSGPT 1663
 Db 578 YLSRGAGPGMGVGHVTRSLLETVAGLLGVKAGAAVPLEPEYPAQRLRLMLEE--RPV 636
 Qy 1664 IVLIGHDTAPDPDTEVNVFVRINDALNSNADGFEVIEHDSKP-----SATSLAYVLYT 1719
 Db 637 VWL-----NVTSEVMTQPDTPNPNPLATPADLAYVLYT 669
 Qy 1720 SGSTGRPKGYMIEHRVIRIWTVS-GCIPNYPSETRMAHMATIAFGASYEIYSALLFGRT 1778
 Db 670 SGSTGRPKGVQITHQAVVNLSSMRHEPGISDRDTLLALITFMEDISALEIFLPSAGAR 729
 Qy 1779 LVCVDYMTTLDARALKDVFFREHVNAASHVTSQQDVLRLVP-----RRLS 1824
 Db 730 VVANQETAVDGERL-----ARELARSKATMMQATPATWRLLLASGWPGRRLT 778
 Qy 1825 RTLMFFLVVTDSTAPDALDAQGLYGVQCYNGYGPTEGVMSTIYPIDSTESFINGVPI 1884
 Db 779 -----ALCGGEALPRDLRLQRTAALWNLGYPTETTIWSAIQVTTGD---GPVSI 828
 Qy 1885 GRALNNSGAYVDPPEQQLGVIGVMGELVVTGDLGARGYSOK-ALDENRFVHITVNDQTVK 1943
 Db 829 GRPIANTQLVLDROMQAPIGVAGELYIGGAGLARGYLNRPPELSADKFFVANSFDPHGR 888
 Qy 1944 AYRTGDRVRYIGGLIEFTGRMDTPKIRGNRIESAEIEAALLRDSVRDAAVVLQONE 2003
 Db 889 LYRTGDLAR-RQRDGALEYLRIDHQVKIRGFIETGEIEAAVRSHPAVRHAVVTAREND 947
 Qy 2004 DOAPEILGFVV-----AD--HDHSENDKQSANQVEGWQDHFESOMYSDIGEID 2050
 Db 948 AAGYLAAYIVPLADGHRATAAATFDHVRSE-----HVTQWQSVMDTITYEQANPAD 1001
 Qy 2051 PSTIGSDFKGMTYDQSIDFDEMHEWGLGTTTTLHDNRSLGNVLGTEIGTSGMILNLD 2110
 Db 1002 PE---FNIVGWRSSVTGEPIPAABEMREWQSDVDRILASRP-RRVLEIGCGTGLLLFRVA 1057
 Qy 2111 SRLSEYVGLPEPSRAAFVNKATESIPSLA---GKAKVQV--GTATDIGQVDDLHPLV 2165
 Db 1058 PHCSEY-----WATDFSQKALDYIAAHADRTGLANVTRFQAADDAACEIDRSQDAV 1110
 Qy 2166 LNSVIOYFSPSEYLAETADTLHLPLNVQRI--FFGDVRSQATNEHFLAARAHITLGNATK 2224
 Db 1111 LNSVIOYFPGAYLRRVLAEARVVKPGGIVFGVGDVRSPLPLETFYASLEVQORAPASLTR 1170
 Qy 2225 DVQRKMAELEDMEEEELLVPAFTSLKDRPPGLVHEVTEILPKNEAVNELSAVRYAAV 2284
 Db 1171 NEFRQVRSLASQEEELVDPAPFFALREQIPE-IGRIEILPRGRSHNELTRFYQAIL 1229
 Qy 2285 HVRSGLGDELVLPEKDDWIDFOANQLNQSLGDLKSSDAAIMAVSKIPEITAFERQV 2344
 Db 1230 HIGSREAE-----PESDRRCQTAAEIR-----VLTAQPELAFAETIPNARLTAE 1280
 Qy 2345 VASLNSIDEWQLSTIRSSAE--GDSSLSVPDIFRIAGEAGFRVEVSARQWSQ---NGA 2399
 Db 1281 VTWNGDEAPTLTGELDRURQTSFGVDPADLWRMDEDLFYRVAI---DWSSHGPHGR 1336
 Qy 2400 LDVAFHCCSOGRTLVNFPDTHLRLGDLTLNRLQRLQNRRIAEVRELRSLPSYMI 2459
 Db 1337 FDATE--CRAAGPPASRPR-RLLAGP--TYNDPLRAYTYTVVQPLRTHLKEKLPDYM 1391
 Qy 2460 PSNIVLDKMLPNANGVDREL-----SRRKVVVQKQTAAPLPTFPISVEVILCEE 2514
 Db 1392 PTAWVLHLEPMTNGKIDRNALPDPEPSRAH-----AEAFPTPETVEQVLAHIW 1443
 Qy 2515 TEVEGKM-VDTIDHFFNLGHSLLATKILSRIDQRKVRITVKDVFDPHPADLASVIRQ 2573

Db 1444 GEVLGMDGIGVHDHFFDGGHSLDLVTQMIAVRDMLHVEVFRFTVFNAPTVRGFAVAIQD 1503
 Qy 2574 GL 2575
 Db 1504 GV 1505
 RESULT 11
 AAR40227
 ID AAR40227 standard; Protein; 3639 AA.
 XX AAR40227;
 AC AAR40227;
 XX 21-FEB-1994 (first entry)
 DT ACVS.
 DE Delta-(L-alpha-aminoacidipyl)-L-cystinyl-D-valine synthase; ACVS;
 KW beta-lactam; antibiotic; transformed; cephalosporin; vector.
 XX Acremonium chrysogenum.
 OS JP05192162-A.
 PN 03-AUG-1993.
 PD 25-JUL-1991; 91JP-0186222.
 PF 31-JUL-1990; 90JP-0205677.
 PR (TAKE) TAKEDA CHEM IND LTD.
 XX WPI; 1993-277475/35.
 DR N-PSDB; AQA48231.
 PT DNA coding delta-(L-alpha-amino-adipyl) L-cystinyl D-valine
 synthase - for improved productivity of cephalosporin antibiotics
 XX Claim 1; Page 14-27; 69pp; Japanese.
 PS The sequence (AAQ48231) is of a vector which includes the
 CC delta-(L-alpha-aminoacidipyl)-L-cystinyl-D-valine synthase gene.
 CC This sequence was transformed into a host cell to express the ACVS
 CC product. The protein produced (AAR40227) was then used to manufacture
 CC a beta-lactam antibiotic.
 XX Sequence 3639 AA;
 SQ
 Query Match 9.4%; Score 1516; DB 14; Length 3639;
 Best Local Similarity 26.1%; Pred. No. 8.3e-112;
 Matches 564; Conservative 377; Mismatches 912; Indels 310; Gaps 80;
 Qy 2 EYLTAVDGQDLPPTPASFCSH-----GDSPLNSSYBOLFHLXGLDSSRTAIAKPCPTPF 55
 Db 793 KYTRASNGTNGVTNGTAHVNGHAANGHVSDSYVASSLQGGFYHSLKNKELSE----- 844
 Qy 56 QLDMDICNALDKQSAIGHAVYDPTDIDISRFALAWKEIVNQTALPAFAFTSDSGKTSQ 115
 Db 845 -----AYTMQSMI-H--YGVPLKRDII--YQAQWQVQGEHPALR-LRFTWEAEVQMI 890
 Qy 116 VILKDSVFVSMCWSSSSSPD-----EVVRDEAAAAA-----SGPRNRFVLLDMQTKK 165
 Db 891 VDPKSELDRVVDWTDVSSREKQLVLELOQTEDLAKYVHLDKGLMLRLYLIL--LPDSK 948
 Qy 166 COLWYTFESHALVDVTFQORVLSRYFAAYKH-----EKD-----THRPEPT 205
 Db 949 YSCIFSCCHHAIQDQWSPLFLFNHNNHQAAYLDLIVGTASPEQDQAYLLCQQLQSHR----- 1004
 Qy 206 ESSDATDTSQSVSVMSCEDNAVSAATHFWQTHLNDLNASVFFHLSHLMVNPPTTAAE 265
 Db 1005 --DDHLDPEWAEQIGRIEERCMDNAL-----LNEASRYKVP-LADYDQV---REQRQ 1049
 Qy 266 HRITFP-----LSOKALSNSAICRTALSILLSRYTHSDEALFG-AVTEQSLPF 312

Best Local Similarity 25.9%, Pred. No. 1.8e-103;
Matches 496; Conservative 320; Mismatches 812; Indels 287; Gaps 64;

```

QY 264 AEHRITPLSOKALSNAICHTAL-----SILLSRYTHSDEALFGA----- 304
Db 1122 AEHRI-----KTLSDVPRTAAIKWCAGHRVTLHSLVOPAMHKLFLAIGSGPVTVGT 1175
QY 305 -VTEQSLPEDKHYLADGTQYQVAPLRVHCOSNLRSADVMAISYDDRLGLHAFGLRDI 363
Db 1176 VISGRGLPVGIDESVGLYINTLPSIVD-HAGQAGQTIAEALADQIESINRMNSRSAREL 1234
QY 364 RNTGDNGSAACDFTVLVTGSHVNGINGFLQQLITESSHFMCNNRALLH---COME 420
Db 1235 FMVOTGSKRRLPETLL-----VLENYPLLLSEEAALHHEHLCFEK 1276
QY 421 SSGALLVAY-----YDHNVIDSLQTRLLQOQGHILKICLOSP 457
Db 1277 SYSDSKVDYFVAVVAREDDGRFEINLWYAGELEFDGTVDNLLGT-LDALFQQAADFDAP 1335
QY 458 LDLSMAEAVNLMTEDRAEIESWNOSPLE-VODTLIHHEMLKAVSHSPTKTAIQAWGDW 516
Db 1336 VE-----RLELVCAADRRRDEWNRTEFAROTLHAGFEIAERWPDIAIVVSGENRL 1390
QY 517 TYGELDNVSRSLAVHKSGLRAQQAIPYVFEKSKWVIAASMLAVLKSGNAFTLIDPNDP 576
Db 1391 SYRELNERANKIAAFLAARVAPKDEPIGLVWEKSEWMAIILGVWKAGAAYVPIDPSYP 1450
QY 577 PARTAQVVTOTRATVALTSKLHRETQKLVG--RCVV--VDELLQSVASDDFSSITKS 632
Db 1451 DDRIRFIFDQARLVLTDRAYAERLDRVADGPFCEVGVQDLPLEPYPAANPRSAAT-S 1509
QY 633 QDLAYVFTSGTGDPKGIMIEHRA-----PSSCALKFGASLGINSDFRALQFGTHAFGAC 688
Db 1510 TDLAYALYTSGTGRKAVLIERGVVNLHLSLERLFDLSRD-RGDVAULFSNIVFDHF 1568
QY 689 LLEIMTTLINGCVCIPSDDRMNS--IPSFINRYNNMMATPSYMGTFSPEDVPGLAT 746
Db 1569 VEQMTDALLSGQTLVMDSDMSRSDQORLYAYMANAVTYLSGTPSVLSLEYGSIPLSR 1628
QY 747 LVLVGQOMSSVNAIWAPKLO-----LNGYQCSRSSICFASN-----MSTEPNMGRA 796
Db 1629 IDAIGEDFTTPV-----FDKIRSSFGLIINGYGPTE--ISITSHKRLYLKHEP-RLDKS 1680
QY 797 VG-----AHSWVIDPNDINRLPIGAVGELVIESPGIARDYIVPPPPPEKSPFFTDIPSWY 851
Db 1681 IGHVPVANTACYVLP--AMQRPVGVGMGELYIGGIVARGYLNRPELTAERFVSNPQSA 1738
QY 852 PANTFPDGAKYTGDLARYASDGSIVCLGRIDSQVKIRQORVELGAIETHLRQOMPDDL 911
Db 1739 GEKALGNARLYKTGDLVRWLPNGELEYLGRNDMQVKIRQORVELGEVEAILASYPGVTR 1798
QY 912 TIWV---BATKRSQSANSTSLIAFLIGSSYFGNRPSDAHILHDHATKAINIKLEQVLR 967
Db 1799 ALVLAHEYAASAGGASOKYLVAFYL-----SABELPETDILHWMRAKLPOAI-- 1847
QY 968 HSIPSYICMELPRTATKIDRRRLRIMGKDILDKOTQGAIVQOAPAPIPVADTAACL 1027
Db 1848 --VPVRIIRILEIPVTSCKLDVKRL-----PATEFALAGSDVYVQPA-----NOTERDL 1895
QY 1028 HSIWQSLGIDPATVNVGATFELGNGNSITAKMV-NMARSVMGMDLKVSNIIYQHPITLAGI 1086
Db 1896 CRLWSKVTGLGPDRIGRSDDFGSGGDSLRALKQAQSIHAFQNFNVAFAHPNIAAQ 1955
QY 1087 SAVVK-----GDPLSYTLIPKSTHEGVEQSYSGRLWFLDOLDVGSILWLIPIYAVRM 1139
Db 1956 AAFLENASAGAQAULSPAGLAANGAAB-PVSLAQERLLFIDDFVEGTAAYNIPFAFAL 2014
QY 1140 ---RGPVNVYDALRRALAEQRHETLRTTFED-QDGVGVQIVHREKLSSEMKVIDLQGSOL 1195
Db 2015 DSDRAEQ--DAVAAALGVLLQRHAALRTLQGLHDGVRVQNV-LPADEALARFEVRSVY 2072
QY 1196 DFEVLNQBOQTP-----FNLSEAGWRATLLRLGEDDH--ILTIYVHHIISDGWSIDVLR 1249

```

```

Db 2073 PDRSALDAELVADFAHVRFLDQELPLRLARLFAVSGDRRSIVVSLVFFHSCDFGWSWRIFR 2132
QY 1250 RDLNOLYSAALKSDKPLSALTPLPIQYSDFAKQKQDQIFBQ--EKQLNMYWKQKQKSSP 1307
Db 2133 EELQALLLEGATAQI-----LPPLPATYADFSVMQROQLSDORLDAFDYQWRSLAGWOP 2186
QY 1308 AKITPDARPAALLSGDAGCVHVTIDGELYOSLRAFPCNEHTTSFVLLAAAPRAAHYRUTA 1367
Db 2187 LQPLDHARPAQFDYLGREIVFDVATTCQDLRLVLAQTRTSFSSVLLAAAYLTLKAYSG 2246
QY 1368 VEDAVIGTPTANRPELEDILIGCFVANTQCMRINDHHDHDTFTGLINQVKAITYTAFAENED 1427
Db 2247 QSDLIVGTPTANRRHGFODGVIGFEVFNLLALRTTIDGRGLLDLYRLTIGGVLQAQLHED 2306
QY 1428 IPFVVVSALQPGSRDLSSTPLAQLIFAV-----HSQKDLGRFKFQGLS 1472
Db 2307 LPPEQLVKRLGV-EKDSRRHPVQAVFSLANKESVDTCWPMRPHAPDDEGR----- 2357
QY 1473 VPVPSKAYTRFDMFHLFOETDSLKGSVNFADFLFKMETVENNVVVFVFEILRNLQSSRT 1532
Db 2358 -----TTAKFEDLSATVSETAQGLSVNFTVAASLFDASSVDNIVATYRRLQEFECR--- 2407
QY 1533 PVSTPLTDCGIVTLEKLDVLNVKHVDYPPRESSLAD-----VFQYQYSAYPDLSLAVV 1583
Db 2408 ---LAPWAESAAL--SDIRYVEACAAPAPAAIAEERAAETLHAAFEKIAAANGDDLLAVA 2461
QY 1584 DSSCRITYTDELDRQSDILA-GWLRRRSMPAETLVAVFAPRSCETIVAFVGLKLANLAYLP 1642
Db 2462 FGOTRLSYRELDARANRLARLSLASRTLAPGDFVALVDKSEWTVVAILAVWKAGAAVYP 2521
QY 1643 LDVRSPEARVQDILSLSGSTIV-----LIGHTDAPPDIEVTNVEFVFIROALN 1691
Db 2522 IDGYPDERIAFLEDTGARAVIADEASYPRISALAGHAARP-----VLNVH-----A 2569
QY 1692 DSNADGFEVIEHDSHTKPSATSLAYLYTSGTGRPKGMIEHRVIRITVTSGCIPNYPSE 1751
Db 2570 HASASSELLPAPASAAGDDDFAYIYTSGTGRPKAVLVHRNVL-SFAGLQRYFGD 2628
QY 1752 TRMAHMATI-----AFDGASYEISALLFGRTLCVVDYMTTLARALKDVFF-REHVNAA 1805
Db 2629 AGAARQAAILFLSSVVFDFVEQLALSILSGNALLIPESLLFD-----DAFYQRMNQERL 2683
QY 1806 SHVTSSQDQVPLRVPRLSRTLMEFFFLAVVTDSTAPDALD-AOGLYQGVQCYNGVGPTEG 1864
Db 2684 SYISGTPQQLFDLAKLDH--LHAVLVAGEAFRAHFHDKIRAEYRG-PLYNAYGTTE-- 2738
QY 1865 VMSTIYPIDSTESFINGVP-----IGRALNNSGAYVVDPEQQLVGIGVMGELVVTGDLAR 1920
Db 2739 --TTVYNL--VKRFADPAPYRNDLGHAIANTELHILDGELRELPPGGLGEIYMGDCVGA 2794
QY 1921 GYSKD-ALDENRV-----HITVNDQTVKAYTGRVRYRIGDGLIEFFGRMDTQ 1969
Db 2795 GYLNRDELSRERFVANPFRSEDDRLGRND---TLYKTGDLVRRR-HSGELEFFGRNDHQ 2850
QY 1970 FKIRGNRIEAEATEALLRSDSSVDAVVLQOQNEQA--PEILGFVVADHDHSEN 2022
Db 2851 VKINGVRIELGEIEAFAAAPPVGAQCAVVAHRHDSGSGSRVLCYCYVAHNVSSES 2905

```

Search completed: May 30, 2003, 12:44:23
Job time : 162 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2003, 12:41:41 ; Search time 37 Seconds
(without alignments)
2488.225 Million cell updates/sec

Title: US-09-482-788-2
Perfect score: 16128
Sequence: 1 MEYLTAVDGRDLPTTASF.....RVEHLLLEEVSKTFEGLNSSL 3129

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pap.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pap.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pap.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pap.*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pap.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5000	31.0	15281	2 US-08-471-119A-2	Sequence 2, Appli
2	1740.5	10.8	3079	4 US-09-413-814-80	Sequence 80, Appl
3	1738	10.8	3072	4 US-09-413-814-93	Sequence 93, Appl
4	1516	9.4	3712	2 US-08-222-617A-4	Sequence 4, Appli
5	1515	9.4	3665	2 US-08-222-617A-13	Sequence 13, Appl
6	1515	9.4	3712	2 US-08-222-617A-25	Sequence 25, Appl
7	1425.5	8.8	2404	4 US-09-134-001C-3464	Sequence 3464, Ap
8	1392	8.6	2628	4 US-09-413-814-11	Sequence 11, Appl
9	1342.5	8.3	3727	2 US-08-222-617A-27	Sequence 27, Appl
10	1342.5	8.3	3728	2 US-08-222-617A-2	Sequence 2, Appli
11	1341.5	8.3	3666	2 US-08-222-617A-12	Sequence 12, Appl
12	855	5.3	739	3 US-08-510-646B-33	Sequence 33, Appl
13	782.5	4.9	1410	3 US-09-335-409-3	Sequence 3, Appli
14	782.5	4.9	1410	4 US-09-568-102-3	Sequence 3, Appli
15	782.5	4.9	1410	4 US-09-567-969-3	Sequence 3, Appli
16	782.5	4.9	1410	4 US-09-568-480-3	Sequence 3, Appli
17	782.5	4.9	1410	4 US-09-568-486-3	Sequence 3, Appli
18	782.5	4.9	1410	4 US-09-568-472-3	Sequence 3, Appli
19	782.5	4.9	1410	4 US-09-567-899-3	Sequence 3, Appli
20	715.5	4.4	797	2 US-08-222-617A-9	Sequence 9, Appli
21	662	4.1	543	4 US-09-413-814-91	Sequence 91, Appl
22	662	4.1	544	4 US-09-413-814-81	Sequence 81, Appl
23	657.5	4.1	768	2 US-08-222-617A-5	Sequence 5, Appli
24	652.5	4.0	798	2 US-08-222-617A-8	Sequence 8, Appli
25	634.5	3.9	758	2 US-08-222-617A-6	Sequence 6, Appli
26	588.5	3.6	2618	4 US-09-413-814-28	Sequence 28, Appl
27	526.5	3.3	822	2 US-08-222-617A-7	Sequence 7, Appli

Sequence 82, Appl
Sequence 94, Appl
Sequence 42, Appl
Sequence 89, Appl
Sequence 76, Appl
Sequence 44, Appl
Sequence 50, Appl
Sequence 8, Appl
Sequence 22, Appl
Sequence 88, Appl
Sequence 4902, Ap
Sequence 86, Appl
Sequence 82, Appl
Sequence 24, Appl
Sequence 94, Appl
Sequence 18, Appl
Sequence 84, Appl
Sequence 92, Appl

ALIGNMENTS

RESULT 1
US-08-471-119A-2
: Sequence 2, Application US/08471119A
: Patent No. 5827706
: GENERAL INFORMATION:
: APPLICANT: Leitner, Ernst
: APPLICANT: Schneider, Elisabeth
: APPLICANT: Schoergendorfer, Kurt
: APPLICANT: Weber, Gerhard
: TITLE OF INVENTION: Cyclosporin Synthetase
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: NO. 5827706artis Corporation
: STREET: 59 Route 10
: CITY: East Hanover
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07936
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/471,119A
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Kassenoff, Melvyn
: REGISTRATION NUMBER: 26,389
: REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201 503 8474
: TELEFAX: 201 503 8807
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 15281 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Tolypocladium niveum
: STRAIN: ATCC 34921
: US-08-471-119A-2

Query Match 31.0%; Score 5000; DB 2; Length 15281;
Best Local Similarity 32.4%; Pred. No. 0;

Matches 614; Conservative 348; Mismatches 858; Indels 381; Gaps 83;

QY 15 PTPASFCHGSDPLNS- YQLEPHLGLDSSRIEAKPCPTFPOLD-----MI 60
Db 898 PTLAQLASHLSSGAASAAATALERGLTRPDGSPSPRVAT--PEEPFALTQCORAWML 955
QY 61 DCNALDKQSAIGHAVYD-----VPTDIDISRFALAMKEIVNOTPALRAFAFTSDGSKTS 114
Db 956 EC-----QKSADG-ALYNLGRVTRVLGAGVDVAALRAPEGLVERHEALRTFLTRDGHPLQ 1010
QY 115 QVILKDFVSWCMWSSSSP-----DEVVRDEAAAAAGGPCRNFVLLDEM 161
Db 1011 QV--HRRVALEW-----AEPFAMALDEREIVARADE--VRRRAFDFLERGPLL-RVHVWRG 1061
QY 162 QTKKCOLWTFSHALVD---VTQORVLSRVFAAYKHEKDTHRPEPESDSDATDTSQSV 218
Db 1062 EGQPLLTVVVHLVDVWYFALLVRELGLYSALRAGR-PPOLPPPS-----SF 1110
QY 219 SVVSMGCD--NAVSAATHFWOTHLN-----DL-----NASVFHLSHLMVNPNTTT 263
Db 1111 FAAGVSCFSPREAGGAAYWRKALDGAATTATIDLPORARHDASPRCRAHAITLKPRLTG 1170
QY 264 AEHRTFPLSOKALSNSAICRTALSILLSRYTHSDREALFCAVTEQSLPFQKHLYADGTQY 323
Db 1171 ALARLA---RERGTTTLFVLSLTSALTVLLHRASQNDLVVG-VPSAGRNDDESTRAFQYFV 1226
QY 324 TVAPLRVHCOSN-----LRASDV-MDAISSYDDRLGLHAPFLGRLDIRNTGNGSNACD 375
Db 1227 QMLPVRLRGAAFDALVARVDAFLDGLHAGDSALQHL-----LAEPRGAARRGGALFD 1282
QY 376 -----FQTVL-----LVTDGSHVNNNGINGFLQQTITESHFHPCNNRALLLHCQMESSG 423
Db 1283 VAFQFQALPSLPRLAALTGAEDVRIAQGELELTTLADEQAAAEFDLALFAELD-SG 1341
QY 424 ALLVAYYDHNVIDSLQTLRLQFQGLHKLKQSPDLSS--MAEVNLMTEYDRA-EIESW 480
Db 1342 IALRFEYDQOLFDPATIERMAR---HFVLLLESAVEHPGRPLSELRLMSDAERALLDDW 1398
QY 481 NS-----QPLEVODTLIHHEMLKA-VSHSPTKTAIQAGDGDWTYSELNDVSSRLAVHKS 534
Db 1399 SGAAAAQAASAPACVHALFEAHAARQPDATALEFGHQRTYAEIWSLSTWSTELALWLKD 1458
QY 535 LGLRAQQAIPVYFKSKWIVASMLAVLKSNGAFTLIDPNDPPARTAAQVVTOTRATVALT 594
Db 1459 RGV-GPGSVGVGCIERSPRWAAQLAVLKAAGAYASLDPANPPARLAEMLADCAALVLT 1517
QY 595 SKLHRETQKLVCR-CVVVDDELLO--SVSASDDFSLSLKSDQLAVITFSGSDGPKGI 651
Db 1518 SS---QASHKLTAAAPCV---HLVODGACAPSTHPIVLSRPDDDLAYVLFVTSGSGTPKGV 1571
QY 652 MIEHRAFSSCALFKFASLGINSSTRALQFTHAFGACILLEIMTLLINGGCVCIPSDDDRM 711
Db 1572 CVRHASLSRLVSFFQHLHLSRDRWTQIASSGFDASVYEIWTPLACGAALLI-ADDDAL 1630
QY 712 NSIPSPINRYNNMM-----ATPSYMGTFSP-----DVPGLATVLV-VGEOMSSSVNAIW 762
Db 1631 RS-PTAL-----VSWLVAQATLSFMTPLAEACFEQDWTGIALRAMTVG---GDKLHPLR 1682
QY 763 AP-KLQLLNGYQSES-----SSICFASNMSTENNNMGRV-GAHSWVIDPDNDINRLVPIGA 817
Db 1683 RPLPRLFNWYGPTEATVITTVAEVADLGDPE-PLGRPTDSALVYVLDPH--MQPVPGV 1739
QY 818 VGELVIESPGIARDYITVPPPEKSPFFTDIPSWPANTFPD-----GAKLYRTGLDARYA 872
Db 1740 LGELYIGGACLAQY-----RDLTA---ERFLPDFGQPGARLYRTGLDLWR 1786
QY 873 SGGSIVCLGRIDSQVKIRQORVELGAIETHLRQOQPDLLTIVVEATKRSQSANS-LIA 931
Db 1787 PDGOLAFAGRDRDEQVKLRGRRVELGEVESALR-----LPVREGVVLHGGOSAARLIA 1841
QY 932 FLIGSSYFCGNRPSDAIHLHDHATKAINIKLEQVLPRHSPTSYICMELPRATATGKIDRR 991
Db 1842 YVPGCA-----DPPSERDLREGMARLVDPDALVPAHFVLLPALPMSLSGKVDK 1889

QY 992 RLRLMGKDILDKQFGAIVQOAPAPIPVFAD-----TAAKLHSIWQSLGIDPATVNV 1044
Db 1890 LL-----PAPPAAHADYEPPEGELERELAHQSVLHLD-----RV 1925
QY 1045 GA--TFPELGNSTIAIKMV-NMARSVGMMDLKVSNIIYQHTPLAGISAVVKGDDPLSYTL-- 1099
Db 1926 GRHDSFDELGHSLAMQVILGRISSELSGIRTTTLTFLFEHTLHQLAORLSSGAASATAAA 1985
QY 1100 -----IPKSTHGPVEQ-----SYSQRLWFLDQDVGSLWYLPYAVMRGPNVVDAL 1148
Db 1986 ATPASEIAPSLGRAPADEYPLSYEQERLWVLEQLLPGGTAYNVVQAVRLRNLDVVDAL 2045
QY 1149 RRALAALQORHETLRTTFEDODGVQIVHEKISEEMKVIDLCSGLDLPFEV-----LNQ 1203
Db 2046 SSALAALVRHRSLSLRTVFV-ASPTPAQKICEPEAPAEVVDLRTGTPDEAAAAARAWASR 2104
QY 1204 EQTTPFNLSSEAGWRATLLRLGEDDHILTIVMHIIISDGWSIDVLRLQOLYSAALKDS 1263
Db 2105 EQATGFDLARGPVRARLFRDLHDHVCVLVTLSTHIVTDAWSFQPLVRLDLAELYRRRGGG 2164
QY 1264 KDFLSALTPLPLOYSDFAKWKQDFIQE--KQINWKKQKDKSSPAKIPTDFARPALLS 1321
Db 2165 P---ADMPELPLOYVDFAVWQRRHLAGKRLADKLAWHTATLRLGLPVLELQTDPRPPVQT 2221
QY 1322 GDAGCVHTIDGELYQSLRAFCHNEHTTSFVLLAAPRAAHYRLTAVEDAVIGTPIANRN 1381
Db 2222 FGAERVLPLDARLVAQLDELARSRGATRFPMVLLAALGVLLRRSSGQDDLAIGTAVANRP 2281
QY 1382 RPELDTIGCFVNTQCMRINIDHDTFGLTINOVKATTTAAFNEDIPFERVVSALOPGS 1441
Db 2282 RPELEPLGVFFVNTIVMRDLGDPTEFELLSPARKVALEAFHQDPAFFKVVAVNP-R 2340
QY 1442 ROLSSSTPLAOLIPAVHSQK----DLGRFKQGLSVPPVPSKAYTRFEDMEHLEFQETDSLK 1497
Db 2341 ROLSRPLFOVLMVQNAPEALELGEVRIEPLD-LPVEA---TRDLRSVEPRGGRDV 2396
QY 1498 GSNFEADELFKMETVENVVRVFEILLRNGLOSRTPVSIPLD-----GIVLEKLDVL 1552
Db 2397 ISQYNVDLFDAAITDRMLATMOSVLSRATQDPAQVRVLSVAPEDERARALVAMNDTAVA 2456
QY 1553 NVKHVDYPPRESSLADVFQTSAYPDSLAVVDSRCLTYTDELQSDILAGWLRRRSMPA 1612
Db 2457 TPDHL-----RLEEFPERAVEQPDACAVVDAERLLTVELARRAEIAAASRSRGATA 2510
QY 1613 ETLVAVFAPRSCETIVAFVGLKANLAYLPDVRSPSARVQDILSLGSGPTIVLIGHDTA 1672
Db 2511 NALVAVVMEKGWEQVAAVLGLVLRAGAAYLPDPLRPEERLHLL----- 2554
QY 1673 PPDIEVTNVFFVIRDALNDSNAD-----GFEVIEHDSK-----PSATS 1712
Db 2555 -----EHAERVLVLTOSAVDGTIAMPAGIERLAVDADERMREQPVARRPPGGSTDD 2605
QY 1713 LAYVLTSGTGRPKGMIEHRVLIIRTVTS-----GCTPNYPSTETRAHMAHATIAFDGA 1765
Db 2606 LAYVIYTSSTGLPKGMIDHKGAVNTVIDINRRFDVG-----PEDRVLAUSSLSFDLS 2659
QY 1766 SYEIVSALLFGRTLVCVDYMTTLDARALKDVFFREHVNAAHSVTS----- 1810
Db 2660 YVDVEGTLAGAGVAVIPDRTRASDPCGHRELVERERVTVNWSVPALMEMLMDASPCGADP 2719
QY 1811 -----SSQDVPRLVPRRLSRTLMEFFLVVTDSTAPDALDAOGLVQGVQCYNGYGP 1860
Db 2720 ALUSSRLVMGSDWIPKLPDRIRAA-----CRAPRVVSLG-----GA 2757
QY 1861 TENGVMSTIYPIDSTESFINGPIGRALNNSGAYVVDPEQQLVIGVMGELVVTGDLAR 1920
Db 2758 TEASIWSTIAHPADVDPAWRSPYGPPLANQHTYVLDLEGPCTGVPGEIHIGGIGVAL 2817
QY 1921 GY-SKALDENFV-HITVNDQTVKAYRTGDRVVRIGDGLIEFFGRMDTQFKIRNRIE 1978
Db 2818 GYWRDEARTRERFLKHPTTGE---RLYRTGDLGRV-FADGTIELLGRTHQVKIRGRIE 2873

Db 2334 RLSRSLPQVMLVQVNAPTALELGEVRIEPLD-LPVEA---TRFDLRFSEVPERGRDV 2389
Qy 1498 GSVNPADELFKMEIVENVVVFBEILRNGLOSSRTPVSLPLTD-----GIVTLEKLDVL 1552
Db 2390 ISLOYNVDLFAATIDRLMATQSVLSRATQPAQRVRLSVAPEDRERLAVNNDTAVA 2449
Qy 1553 NVKHVDVPRESSLADVQTQVSAYPDSLAVVDSCLRTYIELDQSDILAGLWLRSSMPA 1612
Db 2450 TPDHL-----RLEPPFERAVEQDACAVVDARRUTYELARRAIAAASRGATA 2503
Qy 1613 ETLVAFAPRSCETIVAFGGLKANLAYPLDVEDSPSARVQDILSLGSGPTIVLIGHDTA 1672
Db 2504 NALVAVVMEKWEQVAAVLGVRAGAAYLPDPLPERLRHLL----- 2547
Qy 1673 PPDIEVTVFVRIRDALNSAD-----GFEVIEHDSK-----PSATS 1712
Db 2548 -----EHAERVLTQSAVDCTIAPAGIERLAVDADDERWREQPVARRPPGGSTDD 2598
Qy 1713 LAVLYTSGSGRKGWIEHRVIRIVTS-----GCIPNYPSETRMAHMATIAFDGA 1765
Db 2599 LAVYITSGTGLPKGMIDHRGAVNTVLDINRFDVG-----PEDRVIALSLSLFDLS 2652
Qy 1766 SYEYISALLFGRTVLVCVDYMTTDLARALKDVFREHVNAASHVTS----- 1810
Db 2653 VYDFEGLAAGAVVDPDRTRASDPGHRELVERERVTVNSVFMMLMDASPGAGDP 2712
Qy 1811 -----SSQDPLRVPRRLSRTLMFFLVVTDSTAPDALDAQGLYQGVQCYNGYP 1860
Db 2713 ALLSLRLVMSGDMPIPLDRIARA-----CRAPRVVSLG-----GA 2750
Qy 1861 TENCVMSTIPIIDSTESFINGVPIGRALNNSGAVVDPQOOLVIGVYMGELVVTGDGLAR 1920
Db 2751 TEASIWTAHPADVDPANWSIPYGRPLANOHTVLDEGLEPCPIGVPEHIGGIGVAL 2810
Qy 1921 GY-SDKALDENRFV-HITVNDQTVKAYRTGDRVYRIGDGLIEFFGRMDTQFKIRGNRIE 1978
Db 2811 GYNREARTREFLKHPITGE--RLYTGDLGRY-FADGTIELLGRTDHGVKIRGFRIE 2866
Qy 1979 SAEIEAALLRDSVRDAVVLQONEDQAEILGVV-ADHD 2018
Db 2867 LGETEAAALQHPSEVQAVAAKTDPSGKRLVAVVVGADGD 2907

RESULT 4

US-08-222-617A-4
; Sequence 4, Application US/08222617A
; Patent No. 5882879
; GENERAL INFORMATION:
; APPLICANT: Veenstra, Annemarie E.
; APPLICANT: Martin, Juan F.
; APPLICANT: Garcia, Bruno D.
; APPLICANT: Gutierrez, Santiago
; APPLICANT: Barredo, Jose L.
; APPLICANT: Von Doehren, Hans
; APPLICANT: Palissa, Harriet
; APPLICANT: Van Liempt, Henk
; APPLICANT: Montenegro, Eduardo P.
; TITLE OF INVENTION: A Method for Influencing Beta-Lactam
; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
; TITLE OF INVENTION: Quantities of ActV Synthetase
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McConnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222.617A
; FILING DATE: 04-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; REFERENCE/DOCKET NUMBER: 97,157
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3712 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; NAME/KEY: Modified-site
; LOCATION: 2555
; OTHER INFORMATION:
; OTHER INFORMATION: /note= "Xaa-Ala or Ser"
US-08-222-617A-4

Query Match 9.4%; Score 1516; DB 2; Length 3712;
Best Local Similarity 26.1%; Pred. No. 4e-123;
Matches 564; Conservative 377; Mismatches 912; Indels 310; Gaps 80;
Qy 2 EYLTAVDGRQDLPTPASFCSH-----GDSPLNSSYEQLPHLYGLDSSRIEAIKPCPTPF 55
Db 866 KYTKASNGTNGVTNGTAHVNGHAANGHVSDSYVASSLQQGFEVYHSLAKNELSE----- 917
Qy 56 QLDMDICNALDKQAIGHAVYDVPTDIDISRFALAWKEIVNOTPALRAFAFTSDSGKTSQ 115
Db 918 -----AYTMQSMI-H--YGVPLKRDII--YQAARVQGEHEPALR-LRTWEAEVQMI 963
Qy 116 VILKDSVFESVMCWNSSSPD-----EVRDEAAAA-----SGPRCHRFVLLDMDQTKK 165
Db 964 VDPKSELDRVVDVTDVSSREKQLVALEQLQTEDLAKVYHLDKGPLMLYLIL--LPDSK 1021
Qy 166 COLVMTESHALVDVTFQORVLSRVFAAYKH-----EKD-----THRPTP 205
Db 1022 YSLFSCCHHALDGNLSPLLFNNVHQAYLDLVEGTASPEQDAYLYLGGQYLQSHR----- 1077
Qy 206 ESSDATDTSQSVSVSMSCDNVAVSATHFWQTHLNDLNAVFPPLSHLDMVNPPTTTAE 265
Db 1078 --DDHLDFWAEQIGRIEERCDMNA-----LNEASRYKVP-LADYDQV--REQRQ 1122
Qy 266 HRITFP-----LSOKALNSAICRTALSILLSRYTHSDEALFG-AVTEQSLPF 312
Db 1123 QTISLPWNNSMDAGVREELSSRGITLHSILOTVVHLVLSYGGGTHITGTITISGRHLVP 1182
Qy 313 DKHYLADGTQYTVAPL--RVHCOSNLRASVDMAISSYDDDRGLHGLAPFGLDRIRNTGDN 369
Db 1183 PGTERSGLFINTLPMIFDHTVCOD----MTALGAIEHVQGVQVNMNRSRGVGLGRMSKN 1238
Qy 370 GSAACDFQTVLLVTDGSHVNN-----GINGFLOQITESSHFMPCCNNRALLHC 417
Db 1239 DLKHGLFDTLVLENYPNLDTEQREKHEEKLFKTIKGTESKLSYP-----LAVIAQE 1290
Qy 418 QMESSGALLVAY-----YDHNVIDSLQTRLLQQFGHLIKCLQSPDLSSMAEVNLMTEY 472
Db 1291 DGDSCGFLLCYAGELETFDESIALDLTVR--DTLSIDLGINHAPI-----RNMEYLSN 1343
Qy 473 DRABIESWNSQPLEVQDTLTHHEMLKAVSHSPKTAIQANDGDWTSYSLDNVSSRLAVHI 532
Db 1344 QTAQLDKWNATAFEPYNTTLHAMFESAQQPKDVAVVYEDIRLTYRELNSRANALAPYL 1403
Qy 533 KSLGLRAQQAIIPIYVFEKSKWVIASMLAVLSKSGNAFTLIDNPDPAPARTAAQVVTOTRATVA 592
Db 1404 LSQAIIQPNKLVGLIMDKSEHMITSLAVNKTGGAYVPIDPRYPDQRIQYILEDTAALV 1463
Qy 593 LTSKLHRETQVKLVGR--CVVVDDELQSVASDDFSLTKSQDLAVYFTSGTGDPKG 650
Db 1464 ITDSPHIDRLRSITNNRLPVIQSDALQLPSPVHPVSNCKPXSPLAVIMYTSGTGNPKG 1523
Qy 651 IMIEHR-----AFSSCALKFGASIGINSDTRALQFGTHAFGACALLEMTTLINGCVCIP 705

FEATURE:

NAME/KEY: Protein
LOCATION: 1..3665
OTHER INFORMATION: /label= ACVS
OTHER INFORMATION: /note= "ACV Synthetase from Acromonium
OTHER INFORMATION: chrysoenum; aa 1-3665"
US-08-222-617A-13

Query Match 9.4%; Score 1515; DB 2; Length 3665;
Best Local Similarity 26.1%; Pred. No. 4.8e-123;
Matches 564; Conservative 377; Mismatches 912; Indels 310; Gaps 80;

QY 2 EYLFAVGRQDLPPTPASFCSH-----GDSPLNSSYEQLFHLYGLDSSRIEAIKPCPPF 55
DB 866 KYTKASNGTNGVTNGTAHVNGHAANGVSDSYVASSLQGGFVYHSLKNLSE-----917
QY 56 QLDMDICNALDKQSAIGHAYDVPTDIDISRFALAKWEIVNQTPALRAFAFTSGSKTSQ 115
DB 918 -----AYTMQSMI-H-YGVPLKRDY--YQAAQWVQGEHPALR-LRETWEAEVQI 963
QY 116 VILKDSFVSMCMSSSSSPD-----EVVRDEAAAAA-----SGPRCNRFLLEDQTKK 165
DB 964 VDPKSELDRVVDWTDVSSREKOLVALEQLOTEDEAKVYHLDKPLMLRLYLIL--LPDSK 1021
QY 166 COLVWTSFHALVDTFOQVLSRVFAAYKH-----EKD-----THRPETP 205
DB 1022 YSCLFSCHHAILDGSWPLPLFNWVQAYLDLVEGTASPVQDATYLLGQQYQLOSHR---1077
QY 206 ESSDATDTSQSVSVSMSCEDNAVSATHFWOHLNDLNASVPHLSHDHLMVNPNTTAE 265
DB 1078 --DDHDFWAEQIGRIEERCDMNL-----LNEASRYKP-LADYQV---RERQ 1122
QY 266 HRTTFP-----LSQKALSNAICTALSILLSRYTHSDEALFG-AVTEQSLPF 312
DB 1123 QTSISLPWNMSDAGVREELSSRGITLHSILOTVWHLVLSYGGGTHITGTTSGRHLPV 1182
QY 313 DKHYLDGTVQTVAPL---RVHCQSNLRASDVNDMAISSYDDRLGHLAPFGLDIRNFGDN 369
DB 1183 PGTERSGLFINTLPIDFTHTVQD---MTALEATEHVQGVQVNMNSRNGVELGRMSKN 1238
QY 370 GSAACDQFOTVLLVTDGSHVNN-----GINGFLOQITESHSEFPCNRRALLHC 417
DB 1239 DLKHGLFDLFLVLENTPNLDEQREKHEEKLTKTIRGTEKTSYP-----LAVIAQE 1290
QY 418 QMESSGALLVAY-----YDHNVDLSQTTLLQOFGHLIKCLOSLDLSMAEVNLMTEY 472
DB 1291 DGDSGCSFTLCYAGELFTDESIOALLDTR--DTLSIDLGNIIHAPI-----RMEYLSN 1343
QY 473 DRAEIESWNSQPLEVQDTLIIHMLKAVSHSPKTAIQAWGDWMTYSELDNVSRLAVHI 532
DB 1344 QTAQLDKWNAFAEYFNTLHAMFESEAOQKPKDVAVVVEDIRLTYRELNSRANALAEVL 1403
QY 533 KSLGLRAQQAIIIPVYFEKSKWVIASLAVLKSNAFTLLDPNDPPARTAQVVTQTRATVA 592
DB 1404 LSQAATOPKNKVLGIMDKSEHMTSILAVKTKGAGAYVPIDPRYPQORIQYILEDTAALAV 1463
QY 593 LFSKLHRETQVLVGR--CVVVDDELQSVASDDFSSLTQSODLAYVIFTSGTGDPKG 650
DB 1464 ITDSDHIDRLSRITNRLNPVIOQDFALQPPSPVHPVSNCKPSDLAYIMYTSCTGNPKG 1523
QY 651 IMIEHR-----AFSSCALAFGASLGINSOTRALQFOTGHAFGACLLIMTTLINGGCVCIP 705
DB 1524 VWEHGVVNLVCSLRL-FG--LRNTDDEVLSFSYNYVDFHFEVQMTDALLNGQTLVVL 1580
QY 706 SDDDR--MNSIPSFNRYNWNMMATPPSYMGTFSPEDV-PGLATVLVGEQSSSYNATW 762
DB 1581 NDEMGRDKERLYRIETNRVYILSGTPSVISMVEFDRFDHRLRRVDCVGEAFSEPY----1636
QY 763 APKLO-----LLNGYGOSESSICFASNMSTEPNNMRAVAH-----SWIDPNDINR 811
DB 1637 FDKIRETFGLIINGYGPTEVSIITHKRPYPPERRTDKSICQOLDNSTSYVL--NDDMK 1694
QY 812 LVPICAGVELIESGIALDIYVPPPEKSPFTDIPSPYANTPPD-----GAKLY 863

DB 1695 RVPICAVGELYLGGDVARGY-----HNRDPLT--ADRFANPFQTEQERLEGRNARLY 1746
QY 864 RTGDLARYA-----SDGSIVCLGRIDSQVKIRGORVELGAIETHLQQQPDLLTIVVEATK 919
DB 1747 KTGDLVRVHNANGDETEYELGRNDFQVKIRGORIELGEIEAVL--SSYPGIKSQSVLAKD 1805
QY 920 RSOSANSTSLIAFLIGSSYFGNRPSSDAHILDHDAITKAINIKLEQVLPRHSIPSPYICMLE 979
DB 1806 RKNDGQK-----YLVG--YF---VSSAGSL---SAQAIRREMLTSLPDYMYPAQLVPAK 1852
QY 980 LPRATGKIDRRRLRIMGKDILDKQTQCAIVQQAIPAPIPVADTAAKLHLSIWQSLGIDP 1039
DB 1853 FPVTVSGKLDKALPVPDPTVEDD-----IVPRTTEVERILAGIWSELLEIPV 1900
QY 1040 ATVNVGATFFELGNSITAIKWVNA-RSVGMDLKVSNIYOHPTLAGIS-----1087
DB 1901 DRISIDYFFSLGDSGLSKLSTKLSAATPALGAVAVSRNLFPSHTIEALUSQMIIRGSNEVK 1960
QY 1088 --AVVKGD-PLSYTLIPKSTHEGVPESYSGRLWFLDQL-----DVGSLWYLIPIYAVRM 1139
DB 1961 DVAVVKGASLDIPLSP-----AQERLMFIEFHGSHGEDTGA--YVNPLOQL 2006
QY 1140 RGPVNDALRRALAALEORHETLRT--TFEDODGVQVQ-IVHEKLSSEMKVIDL--CGSD 1194
DB 2007 HHVCLLESLEKALRDVSRHEALRTLITRTQSSVHCOKILDAEAOQLFSVDVLRLTSE 2066
QY 1195 LDPEVLNQEQTTTPFNLSSEAGWRATLLRLGEDDHIL---TIVMHIIISDGWSIDVLRD 1251
DB 2067 TEMQGRMAESTAHAFKLEDELP IHVRLYQVVVVDGRTLSFASIVCHHLAFDAWSDVDFQD 2126
QY 1252 LNQLYSALKDSKDPLSALTPLPTQYSDFA-KWOKDOFIOEKOL-NYWKQKLDSSPAK 1309
DB 2127 LDFAVHTKHK--AAANLPTLRVQYKEYAIEHRRALRAEQRHVLADYWLKLSDEASY 2184
QY 1310 IPTDFARPALLSGDAGCVHVTIDGELYQSLRAFONEHTTTSFVLLLAFAAHYRLTAVE 1369
DB 2185 LVDPDRPAQFDYTGNDLQFSTTPETAQLKELAKREGSSLYTVVAAAYFLLLVYVYNQR 2244
QY 1370 DAVIGTPIANRPELEDIIGCFVNTQCMRINIDHDTFTGLINQVKATTATAFENEDIP 1429
DB 2245 DITIGTPVAHRNHPDESVVGVFNLLPLRVNVQSODIHG-LIOAVOKELVDAQIHQDLP 2303
QY 1430 FERVVALQPSRDLSTPLAQLIF-----AVHSOKDLGRFKQGLSVVPVSKAYTR 1482
DB 2304 FOEITKLLHV-QHDPGRHPLQLQAVFNWENVPANVHEEQLLQEKY---PPSPLPSAA--K 2356
QY 1483 FDMFEHLFOETDSLKGSVNFADELFKMETVENVVRVFEELRN-GLOSSRTPVSIPLTD 1541
DB 2357 FDLNVTVKESVNSLVNFNYPSTLSFEETVOGFMETFHLLRLQLAHNKASTSLKLSVED 2416
QY 1542 GIVTLEKLDVLYKHVDYIPRESSLADVFOQVSAYPDLSLAVVSDSSRLTYTEL-DROSDI 1600
DB 2417 GVLNPEP---TNLQPSRDSGNSLHGLFEDIVASTPDRIAIDGTRLSYSELNERANQL 2473
QY 1601 LAGLWRRRSPAEPLVAVFAPRSCETIVAFFGVKLANLAYPLDLDVDRSPSARVODILSGLS 1660
DB 2474 VHLIISASIVADDRILLDDKSIDMVIALLVAKAAYVPLDPTYSQRTLELIESS 2533
QY 1661 GPTVILIGHDTAPPDIETVNTVEFVRI--RDALNDSNADGEVIEHDSKPSATSLAVLY 1718
DB 2534 ARTLIITRKHT-PRGCTVANVSVLSDSPETLACLNQOKENPTTSTQKPS--DLAYVIF 2590
QY 1719 TSGSTGRPKGVMEHRVITRTVTSGCIPNYPSETRMAH-----MATIADFCASEYISALL 1774
DB 2591 TSGTGTGPKGVLEHQSVVQ-LRNSLIERYFGETNGSHAFLSNVYVDFSLBQCLSVL 2649
QY 1775 FORTLVCDVDMTTLDAARALKDVFREHVNAASIVTSSQDVPL-RVPRRLRSLTMFFFLV 1833
DB 2650 GGNKLI-IPPEGLTHEAFYDGRREKLSYLSGTPSVLQOIELSRPLH-----LHMVTAA 2703
QY 1834 VTDSTAPDALDAGLYQGVQCYNGYGTENGVMSTIYPIDSTESFINGVPIGRAL---N 1889

Db 2704 GEEFASPEKMRSFAG-QINNAYGITE-----TTVYNIITP--FKGDAPFTKALCHGIP 2756

QY 1890 NSGAYVDPQOLVGIGVMGELVTDGDIARGY-SDKALDENRFV-----HITVNDOT 1941

Db 2757 GSHVVVLDRLQVFPNAVAGELYLGDCIARGYLNQDALTNERFIPNFPFYEKQASDSRP 2816

QY 1942 VKAYTGRVRYRIGDGLIEFGFMDTQFKTGRNRIESAEIEAALLRDSVYDAAVVLOQ 2001

Db 2817 QRLYKTGDLVRFR-GPHHLEVLGRKDOQVKLGRFRIELSEVRDAVLAISAKVEAAVIPKY 2875

QY 2002 DED 2004

Db 2876 DED 2878

RESULT 6

US-08-222-617A-25

; Sequence 25, Application US/08222617A

; Patent No. 5882879

; GENERAL INFORMATION:

; APPLICANT: Veenstra, Annemarie E.

; APPLICANT: Martin, Juan F.

; APPLICANT: Garcia, Bruno D.

; APPLICANT: Gutierrez, Santiago

; APPLICANT: Barredo, Jose L.

; APPLICANT: Von Doehren, Hans

; APPLICANT: Palissa, Harriet

; APPLICANT: Van Liempt, Henk

; APPLICANT: Montenegro, Eduardo P.

; TITLE OF INVENTION: A Method for Influencing Beta-Lactam

; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large

; TITLE OF INVENTION: Quantities of ACV Synthetase

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESS: McDonnell Boehnen Hulbert & Berghoff

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/222,617A

; FILING DATE: 04-APR-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; REFERENCE/DOCKET NUMBER: 97,157

; INFORMATION FOR SEQ ID NO: 25:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3712 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-222-617A-25

Query Match 9.4%; Score 1515; DB 2; Length 3712;

Best Local Similarity 26.1%; Pred. No. 4.9e-123;

Matches 564; Conservative 377; Mismatches 912; Indels 310; Gaps 80;

QY 2 EYLTAVDGRQDLPPTPASFCSH-----GDSPLNSSYEQLFHYGLDSSRIEAIKPTPF 55

Db 866 KYTRASNGTNGVNTAHVNGHAANGHVSDSVASSLQGGFYVYHSLKNELSE----- 917

QY 56 QLDMDICNALQKQSAIGHAVYDPTDIDISRFALAKWEIVNQTALPAFAFTSDSGKTSQ 115

Db 918 -----AYTQMIMI-H--YGVPLKRDII--YQAAWQRVQGGHPALR-LRFTWEAEVMI 963

QY 116 VILKDSFVFSWMCWSSSSPD-----EVRDEAAAAA-----SGPRCNRFVLLDMDQTK 165

Db 964 VDPKSELDMRVVWDTVDVSREKQVLEALQLOTEDLAKVYHDKGLPMLRMLYLIL--LPDSK 1021

QY 166 COLVWTFSHALVDVTFQQRVLSRVFAAYKH-----EKO-----THRPETP.205

Db 1022 YSCLFSCHHAILDGLWSLPLLFNNVHOAYLDLVEGTASPVQODATYLLGOOYLQSHR---- 1077

QY 206 ESSDATDSDQSVSVMSCEDNAVSAHFHWTNLDNLASVPHLSHDHMLPNPTTTAE 265

Db 1078 --DDHLEWAEQIGRIEERCNMAL-----LNEASRYKVP-LADYDQV---REQRQ 1122

QY 266 HRITFP-----LSOKALSNSAICRTALSTILLSRYTHSDREALFG-AVTEQSLEP 312

Db 1123 QTSILPWNNSMDAGVREELSSRGITLHSLQVFWHLVLSYGGGTHITGTGISGRLHPV 1182

QY 313 DKHYLADGTYTQTVAPL---RVHCQSNLRASDVMDAISYDDRLGLHGLAPEGLDIRTGDN 369

Db 1183 PGIERSVGLFINTLPMIFDHTVCQD---MTALEALEHVQGVQVNAAMNSRGNELGRMSKN 1238

QY 370 GSAACDFQTVLLVTDGSHVNN-----GINGFLOQITESSHFPCNNRALLHC 417

Db 1239 DLKHGLFDLFLVLENYPLNLTQREKHEEKLKFTIKGGTEKLSYP-----LAVIAQE 1290

QY 418 QMESSGALLVAY---YDHNVIDSLQTRLLQQFGHLIKLQSPSLDLSMAEVLNLTVEY 472

Db 1291 DGDSCGFTLCYAGELFTDESQALLDVTVR--DTLSDIILGNHAPI-----RNMEYLSN 1343

QY 473 DRAETESWNSOPLEQVDTLIHHMLKAVSHSPKTAIQAWDGDWITYSELDNVSSRLAVHI 532

Db 1344 QTAQLDKWNATAFEPNTTLHAMFESEAQQKPKVAVVYEDIRLTRELNSRANALAFYL 1403

QY 533 KSLGRQAQAIIPIVFEKSKWIASMLAVKSGNAFTLIDPNDPPARTAAQVTTQTRATVA 592

Db 1404 LSQAAITQPNKLVGLIMDKSEHMITSLAVWKTGGAYVPIDPRYPDQRIQYILEDTAALAV 1463

QY 593 LFSKLHRETVQKLVGR--CVVVDDDELQSVASDDFSSLTQKDLAVIYFTSGSTGDPKG 650

Db 1464 ITDSPHIDRLRSITNRLNLPVIOQDFALQPLPPSPVHPVSNCKPSDSLAIYMTSGTGNPKG 1523

QY 651 IMIEHR-----AFSSCALKEFGASLGINSDFALQFGTHAFACLEIMTTLINGGCVCIP 705

Db 1524 VVVEHGVVNLVCSLCRL-FG--LRNTDDEVILSFSNYVDFHFVEQMTDALLNGQLVVL 1580

QY 706 SDDDR--MNSIPSFNRYNVNMMMATPSYMGTFSPEDV-PGLATLVGEQMSSSYNALW 762

Db 1581 NDEMRGDKERLYRIETNRVTYLSGTPSVISMVYEDFRDRLHRLRVDVCGVEAFSEPY---- 1636

QY 763 APKLO-----LLNGYGOSESSSICFASNMSTEPNMGRAVGAH-----SWIDPNDINR 811

Db 1637 FDKIRETFPGLIINGYGPTEVSTTHHKRYPPEPRTDKSIGCQLDNSTSYVL--NDDMK 1694

QY 812 LVPIGAVALGELVIESPGIARDYIVPPPEKSPFFTDIPSWMYPANTFPD-----GAKLY 863

Db 1695 RVPICAVGELYLGGDGVARGY-----HNRPDLT--ADRFANPFOEQERLEGRNARLY 1746

QY 864 RTGDLARYA----SDGSIVCLGRIDSQVIRGQVRELGAETHLRQOMPDLLTIVVEATK 919

Db 1747 KTGDILVRIHNGANGDEIEYLGNRNDFQVIRGIRGIELGEIYAVL--SSYPIKQSVVYLA 1805

QY 920 RSQSANSTSLIAFLGSSYFGNRPSPDAHLTDHATKAINIKLEQVLPRESIFSYICMLE 979

Db 1806 RNDQOK-----YLVG--YF---VSSAGSL---SAQAIRRFMLTSLDTPVWVAQLPIAK 1852

QY 980 LPRTATGKIDRRRLRMGKDILDKOTQGAIVQQAAPAPIPVAFADTAALKHSIWVQSLGIDP 1039

Db 1853 FPVTVSGKLDKALPVPDDTVEDD-----IVPPTVEVERILAGINSELLEIPV 1900

QY 1040 ATVNVCATPFELGGNSITAIKVMNMA-RVGMMDLKVSNYIQHTPLAGIS----- 1087

Db 1901 DRISIVSDPFLSGGSLKSTKLSFAATRALGAVSVRNLFSHPTIEAESQWIRGSNEYK 1960

QY 1088 --AVVKGD-PLSYTLIPKSTHEGVPQSYSGRLMFLDOL-----DVGSLWLVLIPYAVRM 1139

Db 1961 DVAVVKGASLDIPJSP-----AQERLMFIHEFGHSGEDTGA--YNVPLQUL 2006

```

QY 1140 RGVNVDALRALAALRQORHETLRT--TFEDQGVGVQ-IVHEKLSBEMKVIDL--CGSD 1194
Db 2007 HHDVCLSEKALRDVVSRRHEALRTLTRTKSSVHCQKILDAEEAQKLSVDVRLRTSE 2066
QY 1195 LDPEVLNQOETTFNLSAGWRATLLRGEDDHL--TIVMHHISDGSVDVLRD 1251
Db 2067 TEMGRMAESTAHAFKDEELPHIVRLYQVVRDORTLSFASIVCHHLAFDAMSVDVFRD 2126
QY 1252 LNOYLSAALKSDKPLSALPLPIQYSDFA-KWOKDOFIEQEKOL-NYWKOLKDKSSPAK 1309
Db 2127 LDFAVAVHTKH--AAANLPLTRVQYKEYAIEHRRALRAEQHRLADYWLKRLSDMEASY 2184
QY 1310 IPTDFARPALLSGDAGCVHVTIDGELYQSLRAFCEHNTTSFVLLAAFRAAHYRLTAVE 1369
Db 2185 LVPDRPAPQDYTGNDQLQSTTTPETTAQLKELAKREGSSLYTVVAAAYFLLLYVYNQR 2244
QY 1370 DAVIGTPIANRPELEDDIIGCFVNTOCMRINIDHDTFTGLINOVKATTTAAFENEDIP 1429
Db 2245 DITIGIPVARNHPDFESVVGFFVNNLLPLRVNYSQSDIHG-LIQAVOKELVDAQIHODLP 2303
QY 1430 FERVVSALQSGSRDLSSTPTLAQLIF-----AVHSQKDLGRFKFQGLSVPPVPSKAYTR 1482
Db 2304 FQETTKLLHV--QHDPSRHPQLQAVFNWENPANVHEBQLQYK-----PPSLPSAA--K 2356
QY 1483 FMEFHLFOETDSLKGSVNFADFLFKMETVENVRVFEILRN-GLOSSRTPVSIILPTD 1541
Db 2357 FDLNVTVKESVNSLVNMFNPTSLFEETVQGMETFFHLLRLAHNKASTLSKLSVED 2416
QY 1542 GIIVLEKLDVNLKVDYPRESSADVFOQVSAVPSDLAVDSRCLTYTEL--DROSDI 1600
Db 2417 GVLNPEP--TNLQPSRDSGNSLHGLFEDIVASTPDRIADGTRLSYSELNERANQL 2473
QY 1601 LAGWLRSSMPAETLVAFPAPRCETIVAFGVKLANLAYLPLDVRSPSARVODILSGLS 1660
Db 2474 VHLIISASIVADDRALLDKSIDMVIALLVAKAGAAVPLDPTYPSTQRTLEILESS 2533
QY 1661 GPTIVLIGHDTAPPDIEVTNVEFRI--RDALNDSNADGEVIEHDSKPSATSLAVLY 1718
Db 2534 ARTLIITRKIT--PRGTVANVPVSVLSDPETLACLNOQSKENPTTSTQKPS--DLAYVIF 2590
QY 1719 TSGSTGRPKGMIEHRVIRTVTSGCIPNYPSETRMAH-----MATIAFGASYEIIYSALL 1774
Db 2591 TSCTTGKPKGLVHEQSVQV--LRNSLIERYFGETNGSHAVFLSLNYSVDFSLQCLSLVL 2649
QY 1775 FGTILVCVDYMTILDARALKDVFREHVNAAHVTSSQDVPL-RVPRRLSRLTMEFFLV 1833
Db 2650 GGNKLI--IPPEGLTHEAFYDGRREKLSYLSGTPSVLQQLIELSLPH-----LMHVTA 2703
QY 1834 VTDSTAPDALDAOGLYQGVQYNGYGTENGVMSTIYPTIDSTESFINGVPIGRAL---N 1889
Db 2704 GEEFHASQEKMSQFAG-QINNAYGITE-----TTYVNIITT--FKGDAPFTKALCHGIP 2756
QY 1890 NSGAYVVDPEQVLGVGMELVVTGDGLARY--SDKALDENRFV-----HITVNDQT 1941
Db 2757 GSHVYVNDRLQRPENPANGELYLGGDCLARGYLNQDALNTNREIFPNFEPKQASDRP 2816
QY 1942 VKAYRTGDRVRYRIGDGLTEFFORMTQPKIRGNRIESABIEAALLRDSVRDAVVLQ 2001
Db 2817 QRLYTKGDLVRFER-GPHHLEYLGRKQOVKLRGRIELSEVRDAVLAISAVKEAAVIPKY 2875
QY 2002 NED 2004
Db 2876 DED 2878

```

```

RESULT 7
US-09-134-001C-3464
; Sequence 3464, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

```

```

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3464
; LENGTH: 2404
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-3464

```

```

Query Match      8.8%; Score 1425.5; DB 4; Length 2404;
Best Local Similarity 23.2%; Pred. No. 1.7e-115;
Matches 582; Conservative 466; Mismatches 1032; Indels 431; Gaps 96;

QY 68 QSAIGHAV---YDVPTDIDISRFALAWKEIVNQTPALRAFAFTSDSGKTSQVILKDSFVF 124
Db 36 QSSSEHRTYVWKLRSRDIELRLMYALLDVVQSQPLRT-QFVTTDDFNOLKINLRDFF-- 92
QY 125 SWWCWSSSSPDEVVRDEAAAAAGPCNRNRFVLEDM--QTKKCOLVWTFESHALVDVTFQ 182
Db 93 -----PFIEIKVENMSQSIDLEAFFTRNLNSYHFN 123
QY 183 QRVLSRVFAAYKHEK-----DTHRPEPTESSDATDTSQSVSVVSMSCDNVSAATHFW 236
Db 124 QLPFN-FKIYQFLDEAYLLLDLPHATFNESQLTPFLOQLNIATYHLSKSEYSISDFYNW 182
QY 237 QTHLN---DLNASVPHLSHLMVNP-----TTTAHRITTPLSOKALSNSAIC 283
Db 183 IKEMNQMDONQVQVCP--SRHFNVLNADGONYAYIPVKNTECKKKMCSLHAE-LPSLDID 239
QY 284 RTALSILLSRY--THSDEALFGVTEQSLPDKHYLDG-----TYQTVAPLRVHCQ 333
Db 240 WVIVSYLAHHFISQSSDVTLGI-----HFSIDNKNTENMMVNLNTDIAPLNISIS 289
QY 334 SNLRASDVMDAISYDDRLGLHAPFLGRDIRNTGDN-----GSAACDFQTVLLVTDGSHV 388
Db 290 QSDVWVDMVDECSAL-----LEELQMGASFPVQPKAVQIDVEIMI-----HI 332
QY 389 NNGINGFLOQITESSHPMPCNNRALLLHCOMESSGALLVAYYH-----NVIDS 437
Db 333 EK-----VQSQFELNHI--CHHIRLYN--EASSFADLEFYPHVODGFVDIVNDNYDD 382
QY 438 LQTRLLQQF-GHLIKLQSLPDLSSMAEVLNMTEDYDRAETESWNSOPLVQDTLIHEM 496
Db 383 LTVHTLVKLLINGIYMQITQNPSSL--IKDKLSRSLAKYNDINLQNDININSEVITYKT 440
QY 497 L-----KAVSHSPTKTAIQAWGDWTYSSELDNVSSRLAVHIKSLGLRAQQAIIIPVYPEKS 551
Db 441 VVERFERQVHOHPDSIALQVEQRSMYVHQLNQCANLLAYRLNLHQIETPNMDVIALIERS 500
QY 552 KWTIASMLAVLKSNAFTLIDPNDPARTAQVVTQTRATVATLTKLHRETQVQLVGRCVV 611
Db 501 LEMIICMLGILKAGAGYIPIDPDYPERMNYIITEDAKPAVAVT---YRTSFQSLGPO--- 554
QY 612 VDDELQSVSASDDFSS---LTKSQDLAVYVFTSGSGDPKGMIEHRAESSCALKFGAS 668
Db 555 MDIELIVD-SREHDINDPRINCISEDIAVYVYSGTIGKPKGLVPHRGIDR-LVHNPNY 612
QY 669 LGINSSTRALQFGTHAFGACLLLEIMTTLINGGCVCIPSDDRMNS--IPSFIRNYNNNM 726
Db 613 VELNENTVLLSCTVAFDAATFEIYGPLLNGGRVLITSKDTLLNPLQLDQAITEKNVNTM 672
QY 727 MATPSYMGTFSPEDVGLATL--VINGEQMSSSVNAIWA PKL-----QLLNGYGOSES 777
Db 673 WLTSSLFNQIASERIEALLESLEYLLIGGEV---LNAKVVHLLNSRECHPOIINGYGPTE 729
QY 778 SSICFASNSTENNNGRAV-----GAHSWIDPDNDINRLVPIGAVGELVIESGCIAR 830

```

Db 730 TT--FTTFAI-POEMSPRIPGLPISGTTVVVMOG---DRICGVGVPGELICIGAGLAK 783
QY 831 DYIVPPPE-----KSPFTDIPSWYPANTPPDGAKYRTDGLARYASDSIVCLGRIDS 885
Db 784 GYLQPKLTAEFRFQSPENNM-----LYRSGDLVRLQEDGVIDYISRIDK 829
QY 886 QVKGIRQVELGATETHLRQOMPDLTIVVATKRSQSANSTSLIAFLIGSSYGNRPSD 945
Db 830 QVKGIRFRIELSEKAL-EAIRDINKAVVIREODQ--DKQIVAYEASOL-----878
QY 946 AHILDHATKAINIKLEOVLRPHSPFVYICMLELPRATGKIDRRRLRMKGDKLDKOT 1005
Db 879 -----KSTGOLKDLSETLPEYMPVHFVKVDRIPTMNGKLDVRL-----PELNKANN 928
QY 1006 QGAIVOQAPAPVFPADTAADKHSIWQSLGIDPATVNVGATFFELGNSITAKMNV-M 1064
Db 929 RNYVEPR-----NDIERTVCRIFEILHVD--OVGVKDNFFELGSHSLRATLVNRI 978
QY 1065 ARSVMGDLKVNIVQHPTLAGISAVV---KGDPLSYTLIPKSTHEGPEQSYSGRLWFL 1121
Db 979 EERLKKRLKVGDLKMSPTVEGLGOQIEELQND--VYEVIPKANESYQYDLASOKSMYLL 1036
QY 1122 DOLDVGLWLYLIPVAVMRGPVNVVDALRRALAALEORHETLRTTFEDODGVGVQIVHEKL 1181
Db 1037 WKVNPKDVTYVNIPLFWRLSSELNVQORALKLIERHEILRTQVIDDNEVKORIATHV 1096
QY 1182 SEEMKVIDLGDSDLPFVLNQEOQTPFNLSSEAGWRATLLRLGEDDHILITVMHIIISD 1241
Db 1097 SPDPEEV---TSLTNEQDIIQSEFMEPDLQSPQMRYKYIHGPODY-LFMDTHSIND 1152
QY 1242 GWSIDVLRDLNOLYSAAKSDKPLSALTPLPTQYSDFAKWQKQDOFTEQEKOLNWKYKQ 1301
Db 1153 GMSNTILSDNALYQD-----KSLPELKQYKDYSEWVHR--DLSKQRHFVLOQ 1201
QY 1302 LKDSPP-AKIPDFARPALLSGDAGCVHVTIDGELYQSLRAFPCNEHNTTSFVLLAAPRA 1360
Db 1202 FENQVPLNMTDTPRSIKITNGMLTFHNRQIKQOLKSYVEQHVQVDFWFASALMV 1261
QY 1361 AHYRLTAVEDAVIGTPIANRPELEDIIGCFVNTQCMRINIDHHDTFCTLINQVATTT 1420
Db 1262 LLHKYTRQDDIAIGSVISARTHRTDENMLGMFANTLVYGRPHDQKTDWOLMAEMKML 1321
QY 1421 AAFENEDIPFERVYSALOPGSRDLSTPLAQLIFAV-HSQKDLGRFKFQGLSVPPVPSKA 1479
Db 1322 GAYEHQEPFESLWNL--VDERDASHNPLFDVMLVLQNNETHANFNGHSQTHIP-POST 1379
QY 1480 YTRFDMERHLEQETSLKGSYNFADELFKMETVENVVRVFFELRNGLOSSRTVPSILPL 1539
Db 1380 TAKFDLSLIEEDQDDYVNIENYTDLYKQETIIHAEQLOMIKHVISTENKIQDIDE 1439
QY 1540 TDGIVTLKLDVNLVKHV-----DYPRESSLADVFQTVSAYPDSLAYVDSSCRUTYPEL 1594
Db 1440 NDDL--LWLWD-----KHVNDCSLDLPKNKSIQQLLDHVMKAKADVDALKMNGQSMYQEL 1493
QY 1595 DRQSDILAGWLRRSPAEITLVAVPAPSCETIVAFGVKLANLAYLPLDVRSARVQD 1654
Db 1494 DDYSNSMAQTITQNGERVALTERSFEMVSMIAVLKVGVSYPIDVTPDKRIEF 1553
QY 1655 ILSGLSGGTIVLIGHDTAPPDIEVTVFVRIRDALNSDNGFV-----IEHDSKPS 1709
Db 1554 IIEDEVAVALTYGKAIS-----SHIPVIEDIDNTENKRLNIEYAGNLEDD-----1602
QY 1710 ATSLAYLVLTSGSTRPKGVMEHVRVITRTVTSICI-----PNYPSETRMAHMAITAFDGA 1765
Db 1603 ----MYHIYTSCTGPKKAVSKORNILNV---CAWKRLNLSDDDEVLYQVANYVFDAS 1655
QY 1766 SYEITYSALLFGRLVCVDMYTLDRALKDVFREHVNAAASHVTSSSQDVPLVRPRLSR 1825
Db 1656 ATDFYCSLLNGYPLVIATSVERTNTDLEKLSQENITIAS-----IPLQVYN 1703
QY 1826 TLMFEFL--VYTDSTAPDALDAQGLYQGVQCY-NGYGPTENGVMSTIYPIDSTESFINGV 1882
Db 1704 VMHIFYIPKVIIGAPSTPAFVQHLISKHCDMIVNAYGSPSENTVITTSWIYERKGAIPSTI 1763

QY 1883 FIGRALNNNGAYVVDPEQOLVIGVGMGELVWTGDGLARGYSDK-ALDENRFVHIITVNDQT 1941
Db 1764 PIKGPLANVDIFIMS-GGKLCGVGIPGELCIAGESLTSGLNRPPELSAEKFINNPPGPG- 1821
QY 1942 VKAYRTGDRVYRIGDGLIEFGRMDTOFKRGNRIESAEIEAALLROSSVRDAVVLQO 2001
Db 1822 -OLYRSGGLARL-MPDQIEFLGRIDKQKVHGYRIELGETEININSVDTVTDSVILAK 1879
QY 2002 NEDQAPETILGFVADHDHSENDKQSANQVEGQWQHFESGMYSDIGEIDPSIGSDFKGW 2061
Db 1880 OGER--EVLHAYYVGSQEDENHISOLNAY--LPKYMIPKTLTALISEI-PLT-GND-----1929
QY 2062 TSMWDCSQIDFEDHEHWEIGETTRTLHDNRSLGNVLIEICTGSMI---LFNLDLSRLESYVG 2118
Db 1930 --KVDESRLPVPNVHKNKFPAPRN-NIEREIAQIV-----SGVLDVSSMSIDD-----1974
QY 2119 LEPSRSAAFVFNKATESIPSLAGKAKVQVGTATDQGVDDLDHPLDLVLVNSVIQYFSPSEY 2178
Db 1975 -----DFTFMGCTSLDAMVVVSKL-----KNGIH---ITMQDVYQ-EKTVRY 2013
QY 2179 LAE-----TADTLI--HLPNVQRIFFGDRVSQATNEHFLAARAIHTLGKNAT-----2223
Db 2014 IANTHEKROALPEVVLPDHLPOLOSIV--ERYQLKSOHLTQSSHLGHLVLLTGTGFLGAY 2071
QY 2224 -----KDDVRQKMAELEDKEELLEVPAFFTSKLDRFPGLVHEVILP 2266
Db 2072 LIDEMQDDAQITCIVRGHDINQAKTNLENN-----LNCYPDTAHVDK---LMKHIDIIL 2123
QY 2267 KMEAVNEL---SAYRYAAVHVRSGLDELVPVLEKDDWIDFQANQLNOKSLGOLLKSS 2323
Db 2124 ADLSLDELHIIIDSA--IDTIIHA-GARTDHF-----GDDETFDVFVNRSTQALIDLAKNK 2175
QY 2324 DAAINAVSKIPPEITAFERQVVASLNSNIDEMQNSTIRSSAEGDSSLSVPDIFRIAGEAG 2383
Db 2176 KAKLIYISTI-----SVGTVEVHQDDITTFSEKDIYK--GOL- 2210
QY 2384 FRVESSARQWSQNGALDAVHHCCSQGRTLNVF--PTDHHLRGSDLLTNR 2432
Db 2211 FTSPTYKSKFYSEIKVLEAVNEGLAAQIIRLGNLTSASTGPLNKNLTNR 2261

RESULT 8

US-09-413-814-11
; Sequence 11, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; FILE REFERENCE: PC/US 99/23535
; CURRENT FILING DATE: 1999-10-07
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 11
; LENGTH: 2628
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-11

Query Match 86%; Score 1392; DB 4; Length 2628;
Best Local Similarity 25.3%; Pred..No. 1.9e-112;
Matches 562; Conservative 341; Mismatches 843; Indels 474

Qy	1112	SYSGRLWFLDQDVGSLWLIPIYAVMRGPNVDALRRALAAEQRHETLRTTFEDQDG	1171
Db	24	SFAQRRLWFLDQLEPQSGAVNMPASRTRGPDVDSLVRVNEIVRRHESLRTTVDVIG	83
Qy	1172	VGVOIVHEKLEEMKVTDLCGSLDPE-----VLNQOTTFFENLSSEAGWRATLLRL	1224
Db	84	EPVOVIAPSURIEVPVVDL--SEIDEPEREAARLMAEESRRPFDLTRGPLLRAKLRL	141
Qy	1225	GEDDHILTIYMHIIISDGWSIDVLRDLNOLYSAAKDKSDPLSALTPLOYDFAKWQ	1284
Db	142	GEADHVLLTMHHIVSDGWSMDVLFKELSTLYAAFHGERSP---LEPLQYADFVWQ	198
Qy	1285	KD-QFTEQEQKOLNWKXKQKDS-SPAKIPTDFARPALLSGDAGCVHWTIDGELYQSLRA	1341
Db	199	RELQGEVLESHLYGWEHURGAPTLLELPMDRPRPQAOTFRGSQRAFLPLSLQQAVQA	258
Qy	1342	FCNEHNTSPVWLLAFAFRAAHYRLTAVEDAVIGTPIANRNPELEDIIGCFVNTQCMRIN	1401
Db	259	LSRQEGATPEMTLLTAFSVLLSRYARQSDLVVGTPIANRTRAELEGLEGFVNMALURID	318
Qy	1402	IDHHDFTGTLINOVKATTTAAFNEDIPFERVVSALOPGSRDLSSTPLAQILFAVH---	1457
Db	319	LGGDPSFRELLGVRREVTGLGAYAQHDLPPERLVEELSPG-RSPSHSPLFQVSTLQNTPM	377
Qy	1458	---SOKDLGRFKQGLSVPPVKAYTRFDMEFHLQETDSLKGSYNFADBLPKMEIVEN	1514
Db	378	DATNRADIAS-----GGAPLVEMKAAKFDLLELSESPGGLLTGFYNTDLDAGTIER	431
Qy	1515	VVRVFFELNGLQSSRTPVSIPLTDGIVLEKLDLVNKHVDYPRESSLADVFOQVS	1574
Db	432	MAGHLEVLLSSAAVAPDRPAELPLMGAERSRVLVENWSTAALYPEDHCMHELFEQOVE	491
Qy	1575	AYPDSLAVVDSSCRLTYTDLDRSDILAGWLRSSPAETLVAVFAPRSCETIVAFGVL	1634
Db	492	RSPEAVAVLQOQTLTYRELNMANOLAHRLSLGVGPVRVGLYLERSTETVAIILGVL	551
Qy	1635	KANLAYLPLDVRSPSARVQDILSGLSOPTVLVIGH--DTAPPDIEVTNVFVRIRDALN	1691
Db	552	KAGGAVVLPDPTYPSERL-GLMMADAAPSULLTQASILLSKLPHGDATLVQLOALHEALS	610
Qy	1692	DSNADGFEVIEHDSUTKPSAT--SLAVLYTSGTGRPKGMWIBHRVI--IRTVTSGCIPN	1747
Db	611	R-----LPHETPSGTQAQNLAVMXTSGTGRPKGLVHVRHGLCNLPTVQAKLYGI	662
Qy	1748	YPSETRMAHMATIAFGASVEIYSALLFGRTLVCVDMYMTTLDARALKDVFFREHVA---	1804
Db	663	APGD-RLLOFAPLCFTDSCEIALALLSGATLVWGTADELLPGPPLVELLKKHAYTAMLL	721
Qy	1805	-----ASHVTSSQDVPLRVRRLSRLTMFFVLVVTSDTAPDALDAQGLYQGVQCYNGY	1859
Db	722	APTVLAALPEQOSAAPLRV-----LTMAGEACPAELVVRKWKAPGRRFLNSYSG	769
Qy	1860	PTENGWMSYTIYPIDSTESFINGVPIGRALNNSGAYVVDPEQQLVGVGMGELVVTGDGLA	1919
Db	770	PIETTITWASADLSDERI---PPIGRPLANTQIYVLDEALEPVGIVGPFVIGGVGA	826
Qy	1920	RGYSDKA-LDENRFVHITVND-QTKVAYTRGDRVRVRIKDGLEFFRGMDOFKIRGNRI	1977
Db	827	RGYHGRPDLTAERFVDPDPFGQTKGARLYRTGDRARW-LPDGNLEFLGRNDEQVKVGRVRI	885
Qy	1978	ESAEIEAALLRDSVRDAAVVLOQNEDAQEILGFFVVAHDHSDENDKQGSANQVEGWQDH	2037
Db	886	ELEETRAALLKHPAAQAVAVVRETPGDKRLVAVV-----GRGGARVTA-----	931
Qy	2038	FESGMYSDIGEIDPSTIGSGKWTSMYDGSOIDFDEHMEWLGETTRTLHDNRLSGNVLE	2097
Db	932	--AELRQSVSRILPAT-----WVPSSFALDA-----LPLTPNGKVDKRAUP----	971
Qy	2098	ICTGSGMILFNLDLSRLESYVGLPERSAAAFVNKATATESIPSLAGKAKVGQVGTATIGQVD	2157

[illegible]

Db 1865 EAELEHEHSPLEVEQAH-SOVPRGTPLFESLVFEN--YPVQVTEAPPVEGPTRAEG 1921

QY 3066 VLTKVNIIE-MDEPLYDLAIAAGEPFDGAGLKVTVIAKTQLPGRKRVRHLLFEVSKTFEG 3124

Db 1922 L--RMIDAQISDPYPLTVVAAPH--GTLNLYGERRRDDQAVERMIGHVITLLRG 1976

RESULT 9

US-08-222-617A-27

Sequence 27, Application US/08222617A

Patent No. 5882879

GENERAL INFORMATION:

APPLICANT: Veenstra, Annemarie E.

APPLICANT: Martin, Juan F.

APPLICANT: Garcia, Bruno D.

APPLICANT: Gutierrez, Santiago

APPLICANT: Barredo, Jose L.

APPLICANT: Von Doehren, Hans

APPLICANT: Palissa, Harriet

APPLICANT: Van Liempt, Henk

APPLICANT: Montenegro, Eduardo P.

TITLE OF INVENTION: A Method for Influencing Beta-Lactam

TITLE OF INVENTION: Antibiotic Production and for Isolation of Large

TITLE OF INVENTION: Quantities of ACV Synthetase

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff

STREET: 300 South Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/222,617A

FILING DATE: 04-APR-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

REFERENCE/DOCKET NUMBER: 97,157

INFORMATION FOR SEQ ID NO: 27:

SEQUENCE CHARACTERISTICS:

LENGTH: 3727 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-222-617A-27

Query Match 8.3%; Score 1342.5; DB 2; Length 3727;

Best Local Similarity 21.2%; Pred. No. 9e-108;

Matches 730; Conservative 515; Mismatches 1232; Indels 963; Gaps 134;

QY 312 FDKHYL----ADGTVQTVAPLRVHCQSNLRASD-----VMDAISYDDRLGHILAPGLRDI 363

Db 141 FSVHQLKGFNGGTHITIASL--HREQNLQNSPSSVWYVSPTITVTHENRDGWSVAQAVESI 198

QY 364 ---RNTGNGSAACDFQVLLVTDGSHVNNNGINGFLQQTITSSHF-MPCNNRALLHCOM 419

Db 199 EAARGSESVTAID-----SASSLVKMGFLFDLLVSVFVDDADARIPCFDFPL----- 245

QY 420 ESSGALLVAYYDHN-----IDSLQTTLLRLLQFGHLKICQSPLE---DLSSMAEVLNMT 470

Db 246 ---AVIVRECDANLSLRSDCLFNEETICNFTDALNILLAEAVIGRVTPVADIELLS 301

QY 471 EYDRAIESWNSQPLEVQDT-LIHHEMLKAVHSPSTKTAIOAWDGDWTYSSELDNVSSRLA 529

Db 302 AEQKQLEWNNNTDGEYSSKRLHLHIEEVVERHERDKIAVCDERELTYGELNAQGNLSA 361

QY 530 VHKSIGLRAQQAIIPIVYFEKSKWVIASMLAVLKSNAFTLLIDPNDNPARTAAVYQTIRA 589

Db 1865 EAELEHEHSPLEVEQAH-SOVPRGTPLFESLVFEN--YPVQVTEAPPVEGPTRAEG 1921

QY 3066 VLTKVNIIE-MDEPLYDLAIAAGEPFDGAGLKVTVIAKTQLPGRKRVRHLLFEVSKTFEG 3124

Db 1922 L--RMIDAQISDPYPLTVVAAPH--GTLNLYGERRRDDQAVERMIGHVITLLRG 1976

Db 362 RYLSIGILPEQ-LVALFLDKSEKLIIVTILGYMKSGAAVYPIIDPTYDPDERVRFVLDLTKA 420

QY 590 TVALTSKLRHRETVQKLV---GRCVVVDDDELLOSYSASD-----DFSSSTFKSQDLAY 637

Db 421 RAIILASNGHVERLQREVIGDRNLCLIRLEPLASLAQDSKSKFPAHNLDDLPLT-SOQLAY 479

QY 638 VIFTSGSTGDPKGIIMIEHRAFSKCALKEGASIGINSIDTRALQFGTH-----AFGACLLF- 691

Db 480 VTYTSGTTFGPKGIFKQHTNVVNSITDLSARYGV-----AGQHHEAILLFSACVPEP 531

QY 692 ----TWTLLINGCCVCIPISDDDR--MNSIPSPINRYNNVNMWMTYSYMGTFSPEDVPGLA 745

Db 532 FVRQTLMALVNGHLLAVINDVEKYDADTLLPIRRHSITYLNGTASVLOEYDFSCPSLN 591

QY 746 TLVLVGEOMSSSVNAIWAP--KLQLLNGYQGOSESSI-----CFASNMSTPEPNMGRV-G 798

Db 592 RIILVGENLTEARYLALRQRFKRNILNEYGFTEFAVFTALKIFDEPSTFKDTSLSRGPVRN 651

QY 799 AHSWIDPNDINRLVPIGAVGELVIESPGIARDYIVPP--PPEK---SPFFTDIPSWYPA 853

Db 652 VKCYILNPS--LKRVPIGATGELHIGGLISKGYLNRPELTPHRTIPNPFQDCEKQLGI 709

QY 854 NTFPGAKLYRTGDLARYASDGSIVCLGRIDSQVKIRGORVELGAETHLRROOMDDLTI 913

Db 710 NSL-----MYKTGLARMLPNGEVEYLGRADFQIKLRIGRIEPEGEIETMLAMYPRVTSLS 764

QY 914 VVEATKRSQSANSTS--LIAFLIGSSYFGNRPNSDAHILDHDATKAINIKLEQVLPKHSIP 971

Db 765 VYSKKLRNGPEETTNEHLGVYVCDL---ASVSEADLLSF-----LEKKLPRYMP 812

QY 972 SFYICMLELPTATCKIDRRRLRINKGKIDLKOTOGAIVQOAPAPIPVFADPAALHSTW 1031

Db 813 TRLVQLSQIPVNVNGKADLRALPAVD-----ISNTEVSRLDGRDTEIALGETW 861

QY 1032 VQSLGIDPATVNVGATFFELGCGNSITAIKMNWARS---VGMDLKVSNIYQHTLAGISA 1088

Db 862 ADVLGARQRSVSRNDFRFLGCHSITCQLIARIQRQKLSVISIVEDFATRTLERMAD 921

QY 1089 VVKGDPVLSYTLIPKSTHEGP---VEQSYSGQLWFLDLDLDVGLSW-----YLP 1134

Db 922 LLQKQOEKCDKP---HEAPTELEBENATDNIYLSNQGGFVHYLKSMEQSDAYVWQ 978

QY 1135 YAVRMGPVNVDALRRALAALEQRHETLRTFE-----DQD-----GVGV 1174

Db 979 SVLYRYNTTILSPDLFORANKHAQSPALRLRSEWEKEVQLLDQDPPDLWRELYFTDVA 1038

QY 1175 QIVHEKLSPEMKVIDLGGSDLPFVNLQEQITPNNLSSEACWRATLLRGLGDDHLLTIV 1234

Db 1039 GAVEDRKLEDLR-----RODLTERFKLDVGRUFRVYLIKHSENRRFTCLFS 1083

QY 1235 MHIITSDGWSIDVLRRLNOLYSAALKDQSKDPLSALTPLPIQYSDFAKWKQDQFDEQEQ 1294

Db 1084 CHHAILDGWSLPLLEKVEHYTYLQLLHG--DNLTSSMDP-----YTRQRYLHAHREDH 1136

QY 1295 LNYWKQKLDSSPAKIPDTPDFARPALLS-----GDAGCV-----HVITI--DGELY-OS 1338

Db 1137 LDFWAGVWQ-----KINERCDMALLNERSRYKVQLADYDOVEQRHVTIALSGAWLAD 1191

QY 1339 LRAFCEHNTTSFVVVLLAAAFRAAHYRLTAVEDAVITGPIANRNPDL--EDIIGCFVNTQ 1396

Db 1192 LRQTSQAQGITLHLSILQFVWHAHLAYGGGTHITGTTISGRNLPILGIERAVGPYINT- 1250

QY 1397 CMRINIDH----HDTFGTLINQVKAATTTAAFENEDIPFERV-VSALQPGSRD-----LSST 1447

Db 1251 -LPLVLDHSTFKDKTITIMEAIEDVQAKVNMNSRGNVGLGRHLKHTDLKHGLFSLFVLENY 1309

QY 1448 PLAQLIFAVHSKDLGRKFQGLSEVPVPSKAYTRFDMEFHLFQETDSLKG---SVNPA 1504

Db 1310 PNLDSKRSLEHOTELGYSTEGTEKLNYPYLAIVAR-----EVETGGFTVSICIAS 1360

QY 1505 ELFKMETENVVVRVFFELNGLNGLSSRTPVSLPLTDGTVTLLEKLDLVKLVKHVDYPRESS 1564

Db 1361 ELFEENVISELLHVVQDTLMQVARGNPEVGSLEYLSLI-QUEQLAAMNATEAEP-DTT 1418

GENERAL INFORMATION:
APPLICANT: Veenstra, Annemarie E.
APPLICANT: Martin, Juan F.
APPLICANT: Garcia, Bruno D.
APPLICANT: Gutierrez, Santiago
APPLICANT: Barredo, Jose L.
APPLICANT: Von Doehren, Hans
APPLICANT: Palissa, Harriet
APPLICANT: Van Liempt, Henk
APPLICANT: Montenegro, Eduardo P.
TITLE OF INVENTION: A Method for Influencing Beta-Lactam
TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
TITLE OF INVENTION: Quantities of ACV Synthetase
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,617A
FILING DATE: 04-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 97,157
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3778 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-222-617A-2

Query Match 8.3%; Score 1342.5; DB 2; Length 3778;
Best Local Similarity 21.2%; Pred. No. 9.3e-108;
Matches 730; Conservative 515; Mismatches 1232; Indels 963; Gaps 134;

QY 312 FKHYL-----ADGTYQTAPLRVHCOSNLRASD-----VMDATSSYDDRLGHLAPFGLRDI 363
DB 141 FSVHQLKGFNGTHTITASL--HREQNLQNSPSSWVSGPTIVTHENRDGWSVAQAVESI 198
QY 364 ---RNTGNGSACDFQTVLLVTDGSHVANGINGFLQQTITESHF-MPCNNRALLLHCQM 419
DB 199 EAARGEKESVTAID-----SASSLVKMGFLDLLVSFVDADARIPCFDFPL-----245
QY 420 ESSGALLVAYYDHNV-----IDSLQTRTLQOFGHLIKCLOSPLE---DLSSMAEYVNLMT 470
DB 246 ----AVIVRECDANLSLTFSCLENEETICNFTDALNLLAEAVIGRVTPVADIELLS 301
QY 471 EYDRATEIESNOPLEVDQT-LIHHEMLKAVSHSPKTAIQAWGDWGTSELDNVSSRLA 529
DB 302 AEQKQOLEEWNNTDGEYPSKRLHLHLEEVVERHEDKIAVVCVERELTYGELNAQGNLSA 361
QY 530 VHKSGLGLRAQQAIIIPVPEKSKWVIASMLAVLKSGNATFLIDPNPPARTAAQVVTQTRA 589
DB 362 RYLRSIGTILPEQ-LVALFLDKSEKLIIVTILGVKSGAAAYVPIIDTPYDERVRFVLDTKA 420
QY 590 TVALTSKLHRETQVKLV---GRCVVVDDELQSVASD-----DFSSLTKSODLAY 637
DB 421 RAIASNQHVRLQREIVIGDRNLCIIRLEPLASLAQDSKFFPAHNLDDLPLT-SQOLAY 479
QY 638 VIFTSGTGDPKGMIEHRAFSKALKFGASLGINSIDTRALQFGTH-----AFGACILLE- 691
DB 480 VVTYSGTGTGPKGIFKQHTNVNNSIITDLSARYGV-----AGQHEALLIFSACVFEP 531
QY 692 ----IMTTLINGCVCIPSDDDR--MNSIPSFNRYNVNMNMTATPSYMGTFSPEDVFGLA 745

DB 532 FVRQTLMAVNGHLLAVINDVEKYDADTLLPFFRRHSITYLNGTASVLOQYDFSDCPSLN 591
QY 746 TIVLVGEOMSSSVNAIWAP--KLQLLNGYGOSSESSI---CFASNKSTPNNMGRV-G 798
DB 592 RIILVGENTEARYLALRORFKNRILNEYGFTSFVAFVALKIFDPESTRKDTSLGRPVN 651
QY 799 AHSWVIDPNDINRLVPIGAVGELVIESPGIARDYIVPP--PPEK---SPFTDIPSWYPA 853
DB 652 VKCYILNPS--LKRVPIGATGELHIGLGISKVGLNRPETLPHRFIPNPFQDCEKQLGI 709
QY 854 NTFPDGAKLYRTGDARYASDGSIVCLGRIDSQVKIRGQORVELGALIEHTLRQOMPDLTI 913
DB 710 NSL-----MYKTGDLARWLPNGEYVLGRADFQIKLRGIRIEFEIETMLAMPRVTS 764
QY 914 VVEATKRQSANSTS--LIAFLIGSSVFGNRPNSDAHILDHDKATKAINIKLEQVLPHRIS 971
DB 765 VVSKLRNGPEETNEHLYGYVCDSS---ASVSEADLLSF-----LEKLPYMWIP 812
QY 972 SFYICMLELPTATGKIDRRRLRMKGDIIDKOTQGAIVQOAPAPIPVFADTAALKHSIW 1031
DB 813 TRLVQLSQIPVNVNGKADLRALPAVD-----ISNSTEVRSLRGDTETALGEIW 861
QY 1032 VQSLGIDPATVNVGATFFELGGSITAIKMNWARS---VGMCLKVSNIIYQHTPLAGISA 1088
DB 862 ADVLGARQSRVSRNDFRGLGHSITCICLIARIQRQRSLVSISVEDVFATRLERMAD 921
QY 1089 VVKGDPLSYTLIPKSTHEGP---VEOSYSQGRLWFLDQLDVGSLW-----YLLIP 1134
DB 922 LLQNKQOEKCDKP---HEAPTELLEENATDNIYLANSLQGGFVYHLKSWEOSDAYVMQ 978
QY 1135 YAVMRGPVNDALRALAALAEORHETLRTFE-----DOD-----GVGV 1174
DB 979 SVLRYNTTLPDLFORAKWHAQSFALRLRFSWEKEVFOLQDPLDWRFLYFTDVA 1038
QY 1175 QIVHEKLSSEMKVIDLGGSDLDPEVYNQOETFFNLSSEAGWRATLLRLGEDDHILTIV 1234
DB 1039 GAVEDRKLEDLR-----RQDLTERFKLDVGRFLVFLIKHSENRFTCLFS 1083
QY 1235 MHHIISDGSIDVLRDLNQLYSAAALKDSDPLSALTPLPTQYSDFAKWOKDOFIEQEK 1294
DB 1084 CHHAILDGSLSPLLFKVVHETYLQLLHG--DNLTSSMDPP-----YTRTORYLHAHREDH 1136
QY 1295 LNVKKOLKDDSPAKIPTDFARPALLS-----GDAGCV-----HVTI--DGELY-QS 1338
DB 1137 LDFWAGVYQ-----KINERCDMNALLNERSRYKVQLADYDQVQEQRHVITALLSGDALAD 1191
QY 1339 LRAFCNEHNTSFVLLAAPRAAHYRLTAVEDAVIGTPIANRNPRL-EDIGCFVNTQ 1396
DB 1192 LRQCSAQGITLHSILOFVWHAHLVHAYGGGTHITGTTISGRNLPILIGIRAVGPYINT- 1250
QY 1397 CMREINIDH----HDTFTGLINOVKATTTAAAFENEDIPFERV-VSALQPGSRD-----LSST 1447
DB 1251 -LPVLVDHSTFKDKTIMEALIEDVQAKVNVNVRGNVELGRHLKHTDLKHLGFLDSFLVLENY 1309
QY 1448 PLAQLIPAVHSQKDLGRKFGLEPVPSKATYRDMFHLFOETDSLKG---SVNFAD 1504
DB 1310 PNLDKSRTLHQTELGYISIEGTEKLNYPILAVIAR-----EVETTGGFTVSYCIYAS 1360
QY 1505 ELFKMETVENNVVFFELRLNGLOSSRTPVSYILPTDGIIVTLEKLDVNLKVVHDYDPRESS 1564
DB 1361 ELFEEMWISSELLHMVQDTLMQVARGLNEPVGSLYEYLSLI-OLEOLAANWATEAEFF-DTT 1418
QY 1565 LADVFOQVSAYPDSLAVVDSRCRLTYTELDROSDILAGLRRRSMF-AETLVAVFAPRS 1623
DB 1419 LHEMFENEASQKPKIAVVEETSLTYRELNERANRMAHQLRSDVSPNNEVIALVMDKS 1478
QY 1624 CETIVAFFGVKLANLAYPLDVRSPSARVODILSGLSGPTIVLIGHDTPAPDIEVNTVEF 1683
DB 1479 EHMIVNILAVKMSGGAYVPIDGYPNDRIQVILEDTQ--ALAVIADSCYLPRIKMGMAAS 1536
QY 1684 VRTDALNSNADGFEVIEHDSTKPSATSLAYLYTSGTGKPGKPMIEHRVIRVITGSC 1743

1537 TLLYPSVLPAKDSKWSNSPISLRSTDLAYIIYTSOTGRPKGVTEHHGVNVLQVSL 1596
1744 C-----IPNPSETRMAHMAITAFDCASYEISALLFGRTLVCV-DYMTTLDARALKDVF 1798
1597 SKVFLRDTDDVILS-FSNVDFHVEQMTDAILNGOTLLVNDGM-----RGDKERY 1650
1799 RE-HVNAASHVTSQDQVPLVRPRLSRTLMMFFLVYDSTAPALDAQGLYQGVQYNG 1857
1651 RYIEKNRVTYLSGTPSVVSMYEFGRKDLHRVDCVGEAFSEPVDFDKIRETFHGL-VING 1709
1858 YGPTENGWMS--TIYPIDSTESFINGVPIGRALNAGYVVDPEQQLVGIGVMGELVVTG 1915
1710 YGPTESVITHKRLYPPEPRM---DKSIGQOVHNSYSYVLNEDMKRTPIGAVGELYLG 1766
1916 DGLARGYSKA-LDENRFVHITVNDQTVK-----AYRTGDRVYRIG-DGLIEFFGRM 1966
1767 EGVVGYHNRADVTAERFIPNPFQSEEDKREGRNSRLYKTKDLYRWIPGSSGEVEYLG 1826
1967 DTQFKIRGNRIESAEIAALLRDSVVRDAAVVLOONEDQPEIL-GFVVAD----- 2016
1827 DFQVKIRGLRTEVGEIEAILSSYHGKQSVVIAKDCREGAKQFLVGYVADAAALPSAIR 1886
2017 -----HDSNDKQGSANOVEG----- 2033
1887 RFMQSRLPGYMPVSRLLVSKFPVTPSGKLDTKALPAAEESIDVVPVPRSEIERSLCDI 1946
2034 WOHFES-----GWYS----- 2044
1947 WAELEHMPBEIGIYSDFSSIGGDSLSKLSFMIHESFNRAVSVALFCHRTVEAQTAL 2006
2045 -----DIGETDPSTIGS-----DFKGWTSMDYDGSQIDFDEMHEWLGE 2081
2007 ILNDAADVHETPIDCNDTQMPVSRQAERLLFTHFENGSNAY-----NIDAAFELPGS 2061
2082 TTRTLNDRNSGN-----VLEICTGSGMILFNLS----- 2111
2062 VDSALLEQALGNLARHEALRTLVLKDHATGIYKQVLSPEAQGMFVNVVDTAKQVERL 2121
2112 -----RLES----- 2117
2122 DQETASLSQHVRLDDLPWEARTLKLESGLYLILAFHHTCFDMSLKVFEQELRALYA 2181
2118 GLEPSRAA----- 2126
2182 ALQTKSAANLPALKAKQKEYALYHRRQLSGDRNRNLSDFWLKRLGLEPLQLITDRPR 2241
2127 -----AFVNKATESIPSLAGKAKVQ-----VGTATDIGOVD----- 2157
2242 VQFKYDGGDLSIELSKKETENLRGVAKRCKSSLYVVLVSVVCMVLAANYQNSDVSVGIPV 2301
2158 --DLHPDLVNLNSVQIYFPS-----SEYLAETADTLIHL----- 2189
2302 SHRTHPQ---FQSVIGFVNLVLRVDISOAICGLIRVMKELVDAQHLDQMPFQEVTK 2358
2190 -----PNVORIFFGDRVSQATNEHFLAARATHLTKGN-----ATKDDVR 2228
2359 LLQVNDPDSRHPVQVNF--NFSRANGEDHARSEDGSLAFNQYRPVQVDSVAKEDLN 2416
2229 QKMAEEDMBEELLVEPAFTSL--KDRFPGLVVEHVEILPKNM-----EAVNE--LSAY 2278
2417 ATVTEL---ESGLRVNPNYATSLFNKSTIQGFHLTYEYLLRQLSLSAEGINEDTQLSLV 2473
2279 RYA--AVVHV-----RGSGLDELVLPEKDDWIDF----- 2306
2474 RPTENGDLHLPLAQSPATTAEBOKVASLNQAFERAEFLAAEKIAYVQGDRLSYADLNG 2533
2307 QANQL-----NOKSLGD-----LLKSSDA--AIMAVSK-----IPFEIT----- 2338
2534 QANOLARYIQSVICIGADDGIALMLEKSIDTICILAIWAKAGAYVPLDTPYPPGRVOLLI 2593
2339 --AFEROVVASLNSIDWQ-----LSTRSAGEGSSLSVPDIFRIAGEAGFRV 2386
2594 LEEIKAVLVHSHASKCERHGAKVIAVDSPAETAVSOQSAADLPTIASLGNLA--YI 2651

2387 EVSSAROWSONGAL-----DAVEFH-----C 2407
2652 IFTSGTSKPGKVLVEQKAVLLRDALRERYFGDCTKHGVLFLNSNVFDESVBQLVLS 2711
2408 CSQGRTLNVFTD-----HHLRGS-DLLTNRPLQRLQNRRIAI----- 2444
2712 VLSGHKLIVPAEAFVADDEFYRMASHTGLSYLSGTPSLQKIDLARLDHLQVTAAGEEL 2771
2445 -----EVRER-----LRSLLP---SYMIPSN 2463
2772 HATYEMKRRFNGPIYNAYGVTTVTYVNIITAEFTNSIFENALUREVLPGRVAVVLTAA-- 2829
2464 VVLDKMLNANGKVDRELSRAKVVPKQOATAALPT-----PISVEVEVILCEATEV 2517
2830 -ALQPVFPAVG-----ELYLADGTVTRGYLNQPLLTQDRTIPNPFCKEEDIAMGRFARL 2883
2518 F-----GMKVDTID----- 2526
2884 YKTGDLVRSFRNQOQPQLEYLGRGDLQIKRGYRIEISEVQNVLTSPPGVREGAVVAKY 2943
2527 ---HFFNLGHSLLA-----TKLISRID--ORLVRITVKDFDH-----PVFADLA 2568
2944 ENNDYTSRTAHSLVGYTTDNETVSEADILTFMKARLPTYMPSHLCCLEGALPVTINGK 3003
2569 SVIROGLGLOQVSDGQGDQSAHMAPRTETAILCDEFKVLGFQ-VGITDNFFDLGGH 2627
3004 LDVRR---LPEIINDSA---QSSYSPPRNIIETAKMCRWLWESALGMERCGIDDDLFKLGSD 3057
2628 SLMATKLAVRIGHRLDITVSVKQVDFHPVLEFQALALDNLVQSKTNEIVGGRMAEYSPF 2687
3058 SITSLHLVAQIHNVGCKITVRDIFEH-----RTARALHDHVFMKDS-----DRSNVTQF 3107
2688 QLLFTEDEPEEFMASEIKPQLELQEIQDIYPSQOKAFLEDTHTARTPRPFVFPVYIDPPS 2747
3108 R---TEQGPVIGEAPLLP-----IQDWFLSKALQHPMYNHT-----FVY---R 3145
2748 TSEPDAAGLTKACESLNVNHLIDIPR-TVPAEASGELYQVVLSCLDLPIQVITED-NINTA 2805
3146 TPELDVDSLSAAVARDLQOYHDVFRMLRKREEVGVQSFPAEDFSPAQLRVNLKVDGSA 3205
2806 TNEFLD---EFAKEPVRLGHPILRFTIITKOTKMRVIMRISHALYDGLSLEHVVRKLHM 2861
3206 VNEILLDQWQSGFNLENGPIGS--LOYLHGIEDRSARVWFVHHMAIDTVSQOILVRLQOT 3263
2862 LYNGRSLLP---HQFSRYMQ-YTADGRESHGFWRDVQNTQMTILSDDTVVDGNDAT 2916
3264 LYRNGSLGSGSSPQWAEAIQNYKASDSERNH--WNKLVNEMETASSISALPT-----ST 3315
2917 CKALHLSKIVNIPSOVLGRSSNIITQATVFNACALVLSRESDSKDV-VFGRIVSGROGL 2975
3316 GSRVRLSRSLs-PEKT-----ASLTQGGIDRQDVSVDYSLILTS-VGL 3355
2976 PVEQDIVGCTNAVPVRAH 2995
3356 ALQHIAPTGP---SMWTIEGH 3373

RESULT 11

US-08-222-617A-12
; Sequence 12, Application US/08222617A
; Patent No. 5882879
; GENERAL INFORMATION:
; APPLICANT: Veenstra, Annemarie E.
; APPLICANT: Martin, Juan F.
; APPLICANT: Garcia, Bruno D.
; APPLICANT: Gutierrez, Santiago
; APPLICANT: Barredo, Jose L.
; APPLICANT: Von Doehren, Hans
; APPLICANT: Palissa, Harriet
; APPLICANT: Van Liempt, Henk
; APPLICANT: Montenegro, Eduardo P.
; TITLE OF INVENTION: A Method for Influencing Beta-Lactam

;; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
;; TITLE OF INVENTION: Quantities of ACV Synthetase
;; NUMBER OF SEQUENCES: 27
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: McDonnell Boenhen Hulbert & Berghoff
;; STREET: 300 South Wacker Drive
;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: USA
;; ZIP: 60606
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/222.617A
;; FILING DATE: 04-APR-1994
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; REFERENCE/DOCKET NUMBER: 97,157
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3666 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; HYPOTHEICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Penicillium chrysogenum
;; FEATURE:
;; NAME/KEY: Protein
;; LOCATION: 1..3666
;; OTHER INFORMATION: /label= region
;; OTHER INFORMATION: /note= "Region of ACV Synthetase; aa 62-3727"
US-08-222-617A-12

Query Match 8.3%; Score 1341.5; DB 2; Length 3666;
Best Local Similarity 21.2%; Pred. No. 1.1e-107;
Matches 730; Conservative 515; Mismatches 1233; Indels 963; Gaps 134;
QY 312 FDKHYL----ADGTYQTVAPRVHCOSNLRASD-----VMDAISYDDRLGLHAPFGLRDI 363
DB 80 FSVHQLMGKFGNGTHITIASL--HREQNLQNSPSPVSVPTIVTHENDRGWSVAQAVESI 137
QY 364 ---RNTGNGSAACDFQVLLVTDGSHVNGINGFLQOITESSHF-MPCNNRALLHCOM 419
DB 138 EAARGSEKSVTAID-----SASSLYKMGFLDLLVSFVDADDAIRPCFPPL----- 184
QY 420 ESSGALLVAYYDHNV-----IDSLQTRLLQQFGHLIKCLQSP-----DLSSMAEVLNMT 470
DB 185 ---AVIRECDANLSLTLRFSDCLFNEETICNFTDALNILLAEAVIGRVTPVADIELLS 240
QY 471 EYDRAETESWNSQPLEVQDT-LIHHEMLKAVSHSPTKTAQAWGDWYSELDNVSSRLA 529
DB 241 AEQOQLEEMNNTDGEYSPSSRLHLHIEVYVERHEDKIAVYVCDRELTYGELNAQNSLA 300
QY 530 VHKISGLURAQQAIIPIVYFEKSKVVIASMLAVLKSNGNAFTLIDPNDPAPARTAQVVTOTRA 589
DB 301 RYLSIGILPEQ-LVALFDKSEKLIIVTLGWKSGAAYVPIIDTYPERVRFVLDDTKA 359
QY 590 TVALTSKLHRETVOKLV----GRCVVVDDELQSVSASD-----DFSSITSKSDLAY 637
DB 360 RAIITASNOHVERLOREVIGDRNLCIIRLEPLASLAQDSSKFPFPAHNLDDLPIT-SQQLAY 418
QY 638 VIFTSGSTGDPKGMIEHRAFSKALKFGASLGINSOTRALQFQTH-----AFGACLE- 691
DB 419 VYTSSTGTFGPKGIFKQHTNVNVNSITDLSARVY-----AGQHHEAILLFSACVPEP 470
QY 692 ---IMTTLINGCVCFIPSDDDR--MNSIPSFIRNYNNWMATPSYMGCTSPEDVPCLA 745

DB 471 FVRQTLMALVNGHLLAVINDVEKVDADTLLLPFIRHSITYLNGTASVLOEYDFSDCPSELN 530
QY 746 TLVLVGEOMSSSVNAIWAP--KLQLLNGYGOSESSI----CFASNMTSEPNMGRV-G 798
DB 531 RIILVGENLTARVLAQRKFNRLNEGYFTESAFVTKALFDPESTRKDTSLGRPVN 590
QY 799 AHSWVIDPNDINRLVPICAVGELVIESPGIARDYIVPP--PPEK---SPFFTDIDSWPA 853
DB 591 VKCYILNPS--LKRVPIGATGELHIGGLISKGYLNRPELTPHRIPIPNFQDCEKQLGI 648
QY 854 NTFPDGAKLYRTGDLARYASDGSIVCLGRIDSOVKIRGORVELGAIEIHLRQOMDDLT1 913
DB 649 NSL-----MYKTGLARWLPNGEVEYLGRADFQIKLRIGIRPEGIEIHLWAPVRSLS 703
QY 914 VVEATKRSQSANTS--LIAFLIGSSYFGNRPSSDAIHLHDHDKATKAINKLEOVLPHPHSIP 971
DB 704 VVSKLRNGPEETTNEHLVGVYVCDSD---ASVSEADLLSF-----LEKKLPYRMIP 751
QY 972 SFYICMLELPRATATCKIDRRRLRIMGKIDILDKOTQGAIVQOAPAPIPVPAOTAAALHSIW 1031
DB 752 TRLVQLSQIPVNVNGKADLRALPAVD-----ISNSTEVSRLDGRGDFEIALGETW 800
QY 1032 VOSLGIDPATVNVGATTFEELGNSITAIKVMNMARS---VGMDLKVSNIYQHPTLAGISA 1088
DB 801 ADVLGARQSVSRNDNFELGHSITCQLIARIQORQLSVISVEDVFATRLERWAD 860
QY 1089 VVKGDPLSYTLIPKSTHEGP---VBQSYSGRLWFLDQDLVDGSLW-----YLIP 1134
DB 861 LLQNKQEKCDKP---HEAPTELLBENAATNIYLANSLQQGFVHYLKSMBQSDAYVMQ 917
QY 1135 YAVMRGPVNVDAIRALAALEORHETLRTFE-----DQD-----GVGV 1174
DB 918 SVLYRYNTTSLPDLFORAWKHAQQSPPALRLRFSWEKEVFQLDDQDPLDWRFLYPTDVA 977
QY 1175 QIVHEKLSSEMKVIDLCGSDLPPEVLNBOEOTTPENLSSEAGWRATLLRLGDDHLLTIV 1234
DB 978 GAVEDRKLEDLR-----RQDLTERFKLDVGRFLRVYLKHSNRRFTCLFS 1022
QY 1235 MHIITSDGWSIDVLRDLNLYSAALKDKSKDPLSALTPLPIQYSPFAKWKQDFIQEQK 1294
DB 1023 CHHALDGSPLPLFEKVHETVQLLHG--DNLTSSMDDP-----YTRQRYLHAHREDH 1075
QY 1295 LNYWKKQLKSDSPAKIPTDFARPALLS-----GDAGCV-----HWI--DGELY-QS 1338
DB 1076 LDFWAGVYQ----KINERCDMNAALLNERSRYKVQLADYDQVEQRHVITIALSGDAWLAD 1130
QY 1339 LRAFCEHNTTSFVLLAAFRAAHYRLTAVEDAVITGPIANRNRPCL--EDIIGCFVNTQ 1396
DB 1131 LRQTSAGGITLHSLIQFWHAYLHAYGGGTHITGTTISGRNLPIGLIERAVGPIYNT- 1189
QY 1397 CMRINIDH-----HDTFGTLINQVKATTTAAEFENEDIPFERV-VSALQPGSRD-----LST 1447
DB 1190 -LPLVLDHSTFKDKTITMEAIEDVQAKVNMNSRGNVELCRHLKHTDLKHGLFDSLEFVLENY 1248
QY 1448 PLAQLIFAVHQSQDLGRFKFOGLESPVPVPSKATYTFEDMEHFLFOETDSLKG---SVNRAD 1504
DB 1249 PNLDSRLEHQTELGYSTIEGTEKLNYPYLAIVAR-----EVTGGTFTVCISVAS-1299
QY 1505 ELFKMETVENVRVFEILNRGLSSRTPVSLPLTDGIVTLKEDLVNLKVVDPRESS 1564
DB 1300 ELFEVWISSELLHMVQDTLMQVARGLNEPVGSLVLSSTI-OLEQLAANNATEAEFP-DTT 1357
QY 1565 LADVFQTVQVAYPDSLAVVDDSSCRITYTDLDRQSDILAGWLRRRRSM-P-AETLVAVFAPRS 1623
DB 1358 LHMPENEASQPKDKIAVYVEETSLTYRELNERANMAHOLRSDSPNPNEVIALVMDKS 1417
QY 1624 CETIVAFGVILKANLAYPLDVRSPARVODILSLSGPTIIVLIGHDAPPDIEVTNVEF 1683
DB 1418 EHMIVNIALVWKGSGAYVPIIDPGYNDRIQYILEDTQ--ALAVIADSCYLPRIKMAASG 1475
QY 1684 VRIRDALNDSNADGFEVIEHDSKPSATSLAVLVTSGTSGRPGKGMIEHRRVIRITVTSG 1743
DB 1476 TLLYPSVLPAHPDSDKWSVSNPSLSRSTDLAVIITTSGTGRPKGVTVVEHGVNMLQVSL 1535

QY	1744	C----	IPNYPSETRMAHMTATIAFGASVEIYSALLFGRTLVCV-DYMTTLDARALKDVFV	1799
Db	1536	SKVFLGRD	DDDEVILS-PSNVVDFHFVEQMTDALINGQTLVLVNDGM-----RGDKERLY	1589
QY	1799	RE-HVNAASHV	TSSQDVPLRPVRRLSRTLMFFFLVVTSTAPDALDAQGLYQGVOCYNG	1857
Db	1590	RYIEKNRV	TYLSGTPSVVSMYEFSSRKDHLRRVDCVGEAFSEPVEDKIRFTFHLG-VING	1648
QY	1858	YGPTENG	YMS--TIYPIDSTESFINGVPIGRALNNSGAYVDPDQQLVIGVNGELVWTG	1915
Db	1649	YGPTEV	SLTHKRLYPFERM--DKSIGQQVHNSTSYVLNEDMKRTPICAGELYLGG	1705
QY	1916	DGLARGY	SDKA-LDENRFVHTVNDQTVK-----AYRTGDRVRYIG-DGLIEFFGRM	1966
Db	1706	EGVVG	YHNRAVDTAERFINPQSEDKREGNSRLYTKGDLVRWIPGSSGEVEYLGRN	1765
QY	1967	DTQPK	IRGNIETSAIEAALLRSSVADAAVLQONEDQAPELL-FFVAV-----	2016
Db	1766	DFQVK	IRGLRIELGEIEAILSSYHGIKOSVVIKDCREGAKFLVGYVYADAALPSAATR	1825
QY	2017	-----	-----HDHNSNDKQGSANQVEG-----	2033
Db	1826	RFMQSR	LPGYMWPSRLILVSKFVTPSGKLDTKALPPABEEESIDVVPSEIERSLCDI	1885
QY	2034	WQDH	FES-----GNYS-----	2044
Db	1886	WAELE	HPHEEIGIYSDFFSLGDSLKSTKLSPMIHESFNRAVSVALFCHRTVZEAQTHL	1945
QY	2045	-----	DIGEIDSTIGS-----DFKGWTSMDYDGSQIDFDEMHEWLGE	2081
Db	1946	ILNDA	ADVHEITPIDCNDTQMIPVSRQAERLLFIHEFENGSNAY-----NIDAA	2000
QY	2082	TTTFL	DHNRSLGN-----VLEIGTSGMILFNDS-----	2111
Db	2001	VDASL	LEQALRGMLARHEALRTLVLKDHATGTYLQKVLSPDQAQGMFSYVNDTAKQVERL	2060
QY	2112	-----	-----RLES-----YV	2117
Db	2061	DOETAS	LQHVFRUDDLPWEARILKLESSGLYLILAFHHTCFDAWSLKVFQELRALYA	2120
QY	2118	GLEPS	RSAA-----	2126
Db	2121	ALQKT	SAANLPALKAQYKEYALYHRRQLSGDRMRNLSDFWLKRLKIGLEPLQLITDRPRP	2180
QY	2127	-----	AFVNKATESIPSLAGAKVQ-----VGTATDIGQVD-----	2157
Db	2181	VQKYD	GGDLSLSELSKETENLRGVAKRCKSSLYVLVSVYCVMLASVANDSDVSGIPV	2240
QY	2158	-----	DLHPDLVNLVSVIQYPPS-----SEYLAETADTLHL-----	2189
Db	2241	SHRTH	PQ-----FQSVIGFFVNLVLRVLDISOAICGLIRRVKELVDALQHDQMPQEVTK	2297
QY	2190	-----	PNVQRIFFGQVRSQATNEHFLAARAHITLQKN-----ATKDDVR	2228
Db	2298	LQVND	PDPSRHPVLQNVF--NPESRANGEHIDARSEDEGSLAFNQVRYVPQVDSVAKFDLN	2355
QY	2229	QKMAE	LEDEEELLVEPAFETSL--KDRFPGVLVHEVETILPKNM-----EAVNE--LSAY	2278
Db	2356	AVT	TEL--ESGLRVFNFYATSLFNKSTIQGLFHTYELLRQLSELGSAEGINEDTQLSLV	2412
QY	2279	RYA--	AVVHV-----RGSGLGDELVLVPVEKDDWIDF-----	2306
Db	2413	REP	ENGDLHLPLAQSPPLATTAEQKVASLNQAFREAFLAAREKTAVVQGDRLASYADNG	2472
QY	2307	QANQ	L-----NOKSLGD-----LLKSSDA--AIMAVSK-----IPFEIT--	2338
Db	2473	QANQ	LARYIQVSCIGADGDIALLMEKSIDFTIICILAIWRAGAAVPLDPTYPGRVQLI	2532
QY	2339	-----	AFERQVVASLNSIDEMQ-----LSTIRSSAEGDSSLSVDPDIFRIAGEAGFRV	2386
Db	2533	LEEIK	AKAVLVHSHASKCERHGAKVIVANDSPALETAVSQOOSARLDTIASLGNLA--YI	2590

[illegible]

RESULT 1.2

REC'D 12
US-08-510-646B-33

: Sequence 33, Application US/08510646B

; Patent No. 6077699

GENERAL INFORMATION:

APPLICANT: Blanc, Veronique

APPLICANT: Blanche, Francis

; APPLICANT: Crouzet, Joel

APPLICANT: Jacques, Nathalie

APPLICANT: Lacroix, Patricia

APPLICANT: Thibaut, Denis

APPLICANT: Zagorec, Monique

APPLICANT: Debussche, Lauren

APPLICANT: De Crecy-Lagard, Valerie

TITLE OF INVENTION: Polypeptides Involved In The

; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences

;; TITLE OF INVENTION: Coding For These Polypeptides And Their Use
;; NUMBER OF SEQUENCES: 45
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
;; STREET: 1300 I Street, N.W., Suite 700
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20005-3315
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/510,646B
;; FILING DATE: 03-AUG-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/403,852
;; FILING DATE: 10-MAY-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/FR 93/00923
;; FILING DATE: 25-SEP-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: FR 92/11441
;; FILING DATE: 25-SEP-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meyers, Kenneth J.
;; REGISTRATION NUMBER: 25,146
;; REFERENCE/DOCKET NUMBER: 03806.0054-01000
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 408-4000
;; TELEFAX: (202) 408-4400
;; INFORMATION FOR SEQ ID NO: 33:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 739 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
US-08-510-646B-33

Query Match 5.3%; Score 855; DB 3; Length 739;
Best Local Similarity 31.0%; Pred. No. 4.9e-66;
Matches 229; Conservative 150; Mismatches 314; Indels 46; Gaps 17;

QY 1068 VGMGLKVSNIYGHPTLAGISAVVKGDPVLSYTLIPKSTHEGPEQSYSGRLWFLDOLDVG 1127
DB 5 LGAEATVRDLFEAPTEVAELAELEAREVPPALRAADPEHVPLSPAQRRFLWFLDLEGP 64

QY 1128 SLWYLIPYAVRMGRPVNVVDALRAALAEORHETLRTTFEDQDGVGVQIVHEKLSSEMKV 1187
DB 65 NSTYNIPLALRLRGELDRPALQALTDLTHRHESLRTVPSADGRPYQHVLAPEHAEPLG 124

QY 1188 IDLGSDDLPPFVLNQEQTTPNLSSEAGWRATLLRLGEDDHLLTVMHHIISDGSIDV 1247
DB 125 VVYPADEAGLAEMAAARHEFDVTSEPLRLVSFLTAPDEHVLILLHLLHAGDGWSLAP 184

QY 1248 LRRDLNOLYSAALKSKDPLSALTPLPIQVSDPAKWKQDFTEQ-----KOLNWKK 1300
DB 185 LTRDLTRATAR-RDCAAP--DWEPLPVQYADYTLWQEMLGSPDDPSLGARQLDHWAR 241

QY 1301 QLKDS-SPAKIPTDFARPALLSGDAGCVHVTIDGELYQSLRAFPCNEHNTTSFVYLLAAFR 1359
DB 242 SLAGAPEQLPTDHNRPAAAGHHGRTVPFHEPELHERLSALARSADSLFVLAHAAFA 301

QY 1360 AAHYRLTAVEDAVIGTPIANRNPPELIEDIGCFVNTQCMRINTDHDHDTGTLINQVATT 1419
DB 302 ALLTKHAGTDPIGPIAGRTDEALDDLVGFVNTLVLRDTSGDPTFRELVARTRATD 361

QY 1420 TAAFENEDIPFERVVSALOPGSRDLSSSTPLAQILFAVHSOKDLGRFKFQGLSVVPVPSK- 1478
DB 362 LAAYAHQDLUPFEKLIVETLNP-QKSLARNPLFQVLLAFQSM-PTAQPVLPGLGVVHPEVRV 419

QY 1479 AYTRFDMEFHLFOETD-----SLKGSVNFADFLFKMETVENVVVRVFEILRNGLOSRTP 1533
DB 420 GRAKFDLALAAVEAERHADGRSLRGDWEFSTDLFEQATVEALGARLTALLASVAADPDQP 479

QY 1534 VSILPLTDGIVTLEKLDVLNVKHVDYPR---ESSLADVTQTQVSAYPDSLVAVDSSCLRT 1590
DB 480 IGRVGILD--PAERHRIILTNW-DTSRFGADATWPELFQARAAEHFPAVALVQEGTETG 535

QY 1591 YTDLROSDILAGWLRRRSMPTLTVAVFAPRSCETIVAFVGLVKANLAYLPDVRSPSA 1650
DB 536 YADLNTRANRLARLLRAQIGPEQVVALSPRSADLIVSLAVLTKTGAAYLPVDPAYPAE 595

QY 1651 RVQDILSGISGPTIVLIGHDTA---PPDIEVTNVFVRINDALNSNADGFEVIEHDSTK 1707
DB 596 RIAYLLQD-GAPALVLTHTSVAAGLPGVQLLVQDGLDD-----VPHDILDTDAERTT 648

QY 1708 P-SATSLAYLYTSGTGRPKGVMIEHRVIRVTSGCIP-----NYPSETRMAHMATIA 1761
DB 649 PLHPLPAYVITSGTGLPKGVVPVPH----RSVASVLPVLIBEEFGLGPGSRVLQFASIS 704

QY 1762 FDGASVEIYSALLFGRTLV 1780
DB 705 FDAALWEITLALLSGATLV 723

RESULT 13
US-09-335-409-3
; Sequence 3, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1410
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-335-409-3

Query Match 4.9%; Score 782.5; DB 3; Length 1410;
Best Local Similarity 23.4%; Pred. No. 4.6e-59;
Matches 374; Conservative 247; Mismatches 597; Indels 379; Gaps 64;

QY 1109 VEQSYSGRLWFLDQDVGSLWYLIP---YAVRMGRPVNVDA--LRRALAALAEORHETLR 1163
DB 77 IQESYMWGR-----TGA--FTVPSGHHAYREYDCTDLDVPLSRFRKVVVARHDMRL 126

QY 1164 T-TFEDQDGVGVQIVHEKLSSEMKVIDLCSGLDLPFEV-----LNOEQTTPFN 1210
DB 127 AHTLPDM---MOVIEPKVDADIEIDLRLGDRSTREARLVSLRDAMSHRIYDTERPPL- 181

QY 1211 LSSEAGWRATLLRLGEDDHLLTVMHHIISDGSIDVLRRLDNOLYSAALKD-----SKDP 1366
DB 182 -----YHVAVRDLERQTRLVL-----SIDLNNVDLGSLSIIFKDWLSFVEDP 224

QY 1267 LSALTPLPTQYSDF--AKWKQDFIEOEKOLNWKKOLKDKSSPAKIPTDFAFALLSGDA 1324
DB 225 ETSPLVLEISYRDVYLALESKKSEAHQSRMSDYWKRIAEIPPPPTLPMKADPSTLK-EI 283

QY 1325 GCYVHT--IDGELYQSLRAFPCNEHNTTSFVYLLAAFAAAHRYLTAVEDAVIGTPIANR-- 1380
DB 284 RFRHTEQWLPSDSWGRLLKRRVGERGLTPTGTVILAAAFSEVIGRWSASPRFTLNTLPLNRPL 343

QY 1381 NRPELEIIGCVNTOCMRINIDHDTFTGLINOVKATTTAAAFENEDIPFERY-----V 1434
 Db 344 VHRVNDITGDFTSMVLDDITTRDKSFQRAKRIQEOQLWEAMCHDCVSGIEVOREAARV 403
 QY 1435 SALQPGS--RDLSTPLAQLIFAVHSQKDLGRFKFQGLSEVPVSKAYT-RFDMFHLFO 1491
 Db 404 LGIORGALFPVVLTSALNQVGVTSLSRLG-----TPVYTSQTQOLLDDHOLYE 454
 QY 1492 ETDLSKGSVNADELFKMETVENVVRVFEILR-----NGLOSSPTPVSIPLTDGI 1543
 Db 455 HDGLVLAWDIDVGVFPDLDDMLEAYVFLRLTEPMEQVRCSLPPAQLERASAN 514
 QY 1544 VTEKLDLVNKHVDYIPRESSLADVQTOVSAYPDSLAVDVSSCLRYITELDRSDILAG 1603
 Db 515 ATNALLS-----EHTLGLFAARVEQLPMQJLAIVSARKTLTYEELSRSRRLGA 563
 QY 1604 WLRRRMPAETLVAFPRCETTVAFVGLKAMLAYLPLDVRSPSARVODILSG----- 1658
 Db 564 RLREQGARPNTLVAVMEKGEQVAVLAVLESAAVYPIDADLPAERIHVLLDHGEVKL 623
 QY 1659 -LSGPTVLIGHDTAPDIEVTNVEFVRIKDALNSNADGPEVIEHSTKP-----SAT 1711
 Db 624 VLTQPW--LDGKLSWPEGIQ-----RLVSEAGVEG-----DGDQPPMPIQTPS 666
 QY 1712 SLAYVLTSGTGRPKGMWIEHVRILRTVTS-----GCIPNYPSETRMAHMAITADG 1764
 Db 667 DLAVYVITSGTGLPKGMIDHRAVNTILDINERFEIG-----PGD-RVLAISLSFDL 720
 QY 1765 ASYEIYSALLFRTLVCVDYMTTLDARALKDVFREHVAASHVTSQQDPLRV----- 1819
 Db 721 SVYDVFGLAAGGTIVPVDASKLRDPAHWAELIEREKVTWNVSPAL-----MRMLVEHF 775
 QY 1820 ---PRLSRTLMFFLVTDSTADPAL--DAQGLYQGVQVNGYGPTEGVMSTIYDIDS 1874
 Db 776 EGRPDSLARSURLSL--SGDWIPVGLPGELOAIKRGVSVISLGGATEASIWSGYVPRN 833
 QY 1875 TESFINGVPIGRLNNSGAYVVDPEQOOLGVGMGLVVTGDLGARY--SDKALDENRE 1932
 Db 834 VLSWASIPYGRPLRNTFFHLDLAEPRPVWPGQLYIGVGLALGYWRDEETKRSFL 893
 QY 1933 VHTVNDOTVKAYTGRVRYRIGDGLIEFGRMDTQFKIRGNRIESAIEAALLRDSV 1992
 Db 894 VHPETGE---RLYKTGDLGRY-LPDGNIEFMRGDNQIKLGRYVELGEITEETLKSPNV 949
 QY 1993 RDAVVLQONEDQAPETILGFVADHDHSEKDKGSANQVEGWQHFESGMYSDIGEIDPS 2052
 Db 950 RDAVIVPGNDAANKLLLAYV-----PEGTRRAAEQ-----DASLKE--RIDAR 994
 QY 2053 TIGSDFKGTSMYDGSIDFDEHMEWLGTTRTLHDNRSLGNVLEIGTGSMILFNLSR 2112
 Db 995 AHAAEADG---LSDGERVQFKLARHGL-----RRDL-----DCKPVVDLTGQ 1033
 QY 2113 LESYVGLG---PSRAAFAVNKATESI-----PSLAGKAKV 2145
 Db 1034 DPREAGLDVYARRSVRTLEAPIPFVEFGRFLSCLSSVEPDGATLPKFRYPSAGSYV 1093
 QY 2146 QVGTATDIGOVDLHPLDVLVNSVIOYFPPSEYLAETADTLIHLNPVQRIFFGQVRSQAT 2205
 Db 1094 QTYAYVKSGRIG-----VDEGFYVHPFEHRLKLUSD----- 1126
 QY 2206 NEHFLAARAIHTLCKNAKTDKDVQKMAELEDMEELVEPAFTLSKDRFPGLVEHVEIL 2265
 Db 1127 -----HGTERGA--HVRQNFVDVDEAAFNLL-----FVGRIDALESL 1161
 QY 2266 PKMWEAVNELSAYYAAVHVRVSLGDELVLPEKDDWIDFOANQLNOKSLGDLKSSDA 2325
 Db 1162 -----YGSSSRFECL-----LEAGYMAQ-LLMEQAPSCNI 1190
 QY 2326 AIMAVSKIPFEITAFERQVVASLNSNIDEMQLSTIRSSAEGDSSLSPVDFRITAGEAGR 2385
 Db 1191 GVCVPGQFNPE-----QVRPVLID-----LRHS-----DVI-VHGMILGGR 1223
 QY 2386 VEVSARQWSONGALDAVPHCCSQGRTLVNFPTDHHLRGSDLLTNRLRQLNRRIATIE 2445

Db 1224 VD---PROF-----QVCTLQGD---SSPRATTGGA-----PPGREQH---FADM 1259
 QY 2446 VRERLSLLPSYMIPSNIVWLDKMPLNANGVDKELSRRAKVVPKQOTAAPL---PTFP 2502
 Db 1260 LRDLFTKLPEYMPVPTVEVDALPTNSGKVDKRALRER-----KOTSPRHSGHTAP 1313
 QY 2503 ISEVEVILCEATEYFGMK-VDIITDHFENLGHSLATKLISRIDORLKVRTVYKDVFDH 2561
 Db 1314 RDALKEILVAVREVGLVGLQSQFVDLGATSIHIVMRSLLOKRLDREITAITELFOY 1373
 QY 2562 PVFADLASVIRGGLQLOQPVSDGOGGODRSAMHAPRTE 2598
 Db 1374 PNLGLASGLRR-----DSRDLQORPNMQDRVE 1401

RESULT 14
 US-09-568-102-3
 ; Sequence 3, Application US/09568102
 ; Patent No. 6346404
 ; GENERAL INFORMATION:
 ; APPLICANT: Schupp, Thomas
 ; APPLICANT: Ligon, James
 ; APPLICANT: Molnar, Istvan
 ; APPLICANT: Zirkle, Ross
 ; APPLICANT: Cyr, Devon
 ; APPLICANT: Goerlach, Joern
 ; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
 ; FILE REFERENCE: 4-30582A
 ; CURRENT APPLICATION NUMBER: US/09/568,102
 ; CURRENT FILING DATE: 2000-05-10
 ; PRIOR APPLICATION NUMBER: 09/335,409
 ; PRIOR FILING DATE: 1999-06-17
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 1410
 ; TYPE: PRT
 ; ORGANISM: Sorangium cellulosum
 US-09-568-102-3

Query Match 4.9%; Score 782.5; DB 4; Length 1410;
 Best Local Similarity 23.4%; Pred. No. 4.6e-59;
 Matches 374; Conservative 247; Mismatches 597; Indels 379; Gaps 64;

QY 1109 VEQSYSGRLWFLDQLDVGSWLYLP---YAVMRGPNVDA--LRRALAALEORHETLR 1163
 Db 77 IQESYWLGR-----TGA--FTVPSIHAYREYDCTDLDVPLRSRAFRKVVARHMLR 126
 QY 1164 T-TFEDQGVGVQIVHKEKLEEMKVIDLGGSDLDPFV-----LNQBOTTPFN 1210
 Db 127 AHTLPDM---MQVIEPKVDADIEIDLGLDRSTREARLVSLRDMASHRIYDTERPPL- 181
 QY 1211 LSSPAGWRATLRLGEDDHILITVMHHIISDGSIDVLRLDNLNOLYSALKD---SKDP 1266
 Db 182 -----YHVAVRLDERQTRLV-----SIDLINDLGL-SLIIFKDWLSFYEDP 224
 QY 1267 LSALTPLPIQYSDP--AKWQKQDFIEQEKQLNWKQKQKSSPAKIPDTDFARPALLSGDA 1324
 Db 225 ETSPLVLELSYRDYVLALESKKSEAHQSRMSDYWKRIAEPLPPPTLPMKADPTLK-EI 283
 QY 1325 GCYHVT--IDGELYOSLRACFNEHNTTSFVVLLAAFAAHYRLTAVEDAVIGTPIANR-- 1380
 Db 284 RFRHTEQWLPSSDMSGRLLKRRVGERGLTPTGVILAAFASEVIGRWSASPRFTLNLITLFRNP 343
 QY 1381 NRPELEIIGCVNTOCMRINIDHDTFTGLINOVKATTTAAAFENEDIPFERY-----V 1434
 Db 344 VHRVNDITGDFTSMVLDDITTRDKSFQRAKRIQEOQLWEAMCHDCVSGIEVOREAARV 403
 QY 1435 SALQPGS--RDLSTPLAQLIFAVHSQKDLGRFKFQGLSEVPVSKAYT-RFDMFHLFO 1491
 Db 404 LGIORGALFPVVLTSALNQVGVTSLSRLG-----TPVYTSQTQOLLDDHOLYE 454

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2003, 12:48:42 : Search time 85 seconds
(without alignments)
3726.223 Million cell updates/sec

Title: US-09-482-788-2
Perfect score: 16128
Sequence: 1 MEYLTAVDGRQLPPTPSAF.....RVEHLLVEVSKTEGLNSSL 3129

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 383519 seqs, 10123694 residues
Total number of hits satisfying chosen parameters: 383519

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	1713.5	10.6	2448	10	US-09-815-242-11870	Sequence 11870, A
2	1677	10.4	4342	10	US-09-815-242-5107	Sequence 5107, Ap
3	1632	10.1	4999	9	US-09-976-059-15	Sequence 15, Appl
4	1406	8.7	2397	10	US-09-815-242-12265	Sequence 12265, A
5	1375.5	8.5	4999	9	US-09-976-059-14	Sequence 14, Appl
6	1374	8.5	1668	10	US-09-815-242-5654	Sequence 5654, Ap
7	782.5	4.9	1410	9	US-10-014-717-3	Sequence 3, Appl
8	623	3.9	1051	9	US-09-976-059-13	Sequence 13, Appl
9	612.5	3.8	1446	9	US-10-166-087-44	Sequence 44, Appl
10	559	3.5	1293	10	US-09-815-242-10079	Sequence 10079, A
11	538.5	3.3	1294	10	US-09-815-242-13724	Sequence 13724, A
12	475.5	2.9	891	9	US-09-976-059-18	Sequence 18, Appl
13	463	2.9	458	9	US-10-008-016-4	Sequence 4, Appl
14	434.5	2.7	1391	9	US-09-994-595-8	Sequence 8, Appl
15	424	2.6	1295	9	US-09-738-626-6480	Sequence 6480, Ap
16	415.5	2.6	401	10	US-09-924-256A-88	Sequence 88, Appl
17	390	2.4	485	10	US-09-815-242-5301	Sequence 5301, Ap
18	390	2.4	485	10	US-09-815-242-12515	Sequence 12515, A
19	379.5	2.4	399	10	US-09-924-256A-86	Sequence 86, Appl

Sequence 10589, A
Sequence 82, Appl
Sequence 24, Appl
Sequence 94, Appl
Sequence 11698, A
Sequence 18, Appl
Sequence 42, Appl
Sequence 84, Appl
Sequence 92, Appl
Sequence 13644, A
Sequence 56, Appl
Sequence 3950, Ap
Sequence 90, Appl
Sequence 3817, Ap
Sequence 10, Appl
Sequence 19, Appl
Sequence 19, Appl
Sequence 3, Appl
Sequence 31, Appl
Sequence 24, Appl
Sequence 15, Appl
Sequence 23, Appl
Sequence 16, Appl
Sequence 44, Appl
Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-09-815-242-11870
; Sequence 11870, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11870
; LENGTH: 2448
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11870

Query Match 10.6%; Score 1713.5; DB 10; Length 2448;
Best Local Similarity 28.6%; Pred. No. 2.4e-115;

Matches 586; Conservative 327; Mismatches 750; Indels 385; Gaps 76;

QY	430	YDHNVIDSIQTRLLQOFGCHLIK--CLOQSPDLSSMAEVLNMTXYDRAE--IESWNSQPLE--480
Db	441	YATDLFDASTYERLAGHWNLLRGIVANP--RQRLGELPLDAPERRQOTLSSEWNPQREC 498
QY	487	-VODTLIIHHEMLKAVSHSPTTKTAIQAMDGDWYISELDNVSSRLVAHVHKSGLGQAQALIP 545
Db	499	AVQGTQ--QORFEEQARQPOVALILDFORLSYGENLARANRLAHCLHARGVGAD--VP 554
QY	546	V--YFEKSKWIASMLAVLKSNGNAFTLIDPNDDPAPARTAOVVTQTRATVALTSKLHRETVQ 603
Db	555	VGLALERLMDLWLLATLILKAGGAYPLDPAAPERLAHILDDSGVRLLLTQGHLLERLP 614
QY	604	KLVG--RCVVVDDELQSVASDDFSSLTKSODLAVIFTSGSTGDPKGMIEH---RAF 658
Db	615	RQAGVEVLAIQGLVDGYAESDPLETIS--ADNLAYVITSGTQKPGTLLTHRNAURLF 673
QY	659	SSCALFGASLGINSDFRALQFGTHAFGACILLEMTTLINGGCVCII-----PSDDDR 710
Db	674	SATEAWFGFD---ERDVTWL--FHSYAFDFSWEIEFGALLYGCGLVPIVQWVRSRSPEDFYR 725
QY	711	MNSIPSFNRYNVNMMATPS-----YMGTFSPEDY-----PGLATLVLVEGOMSSSVNAIW 762
Db	730	L-----LCRBGVTVLNOTTSFAKFLMAVACSADMATQOPALRYVIEFGGALDQLSRPW 783
QY	763	APKL-----QLLNGYGQSESSICFASNMSTEP-----NNMGRAVGAHSWIDP 806
Db	784	FORGDRQPOLVNMVIGITEIT-----VHVYIRPVSEADLEGLVSPICGTTIDLSWLID 838
QY	807	NDINRLVPIGAVGELVIESPGIARDYIVPPPPPEKSPFTTIPSWYPANTEPDGA--KLYR 864
Db	839	RDLP--VPRGAVGELYIGRAGLARGYLRRPGLSATRE--VP-----NPPFGGAGERLYR 889
QY	865	TGDLARYSDGSIVCLGRIDSOVKIRQORVELGAEIETHLRQOMDDTLTIVVEATKRS--QS 923
Db	890	TGDLARQADGNIEYIGRIDHVKVGRFIELGEIEAALA-----GLAGYRDAVVLVHADG 944
QY	924	ANSTSLIAFLIGSYFGNRPDSAHILDHIDATKALNIKLEQVLRPHSIPSFYICMLELPT 983
Db	945	VGGTOLGVYVADS-----AEDAERL-----RESLRSLKRLHILFDYVWPAHMLLERPLT 995
QY	984	ATGKIDRRRLRIMGKILDKOTOGAIVQQA--PAPIPVFADTAAKLHSIWQSLGIDPATV 1042
Db	996	VNGKLDRLQAL-----PPDASLSQQAYPAP--GSELEQRIAAIWSEILGVE--RV 1041
QY	1043	NVGATTFELGGSNITAIKMYNMAR--SVGMDLKYSNITQHTLAGISAVYKGDPLSTYLP 1101
Db	1042	GLDDNFELGCHSLLATRVISRVQBOQDLASLKALEFRVLAEFAQGLERTTDAVSTIP 1101
QY	1102	KSTHEGPVEQSYSQGRLWFLDQLDVGSLWYLIPIYAVMRGPVNVDAIARRALAALORHET 1161
Db	1102	LADROQPLALSFAQERQWFLWQLEPESAAHYHPSALRLGRLDVDALQKSFDSLVARHET 1161
QY	1162	LRTF-----EDDQGVGVIVHEKI--SEEMKVIDLCSDDLDPPEVLNQEOTTFFNLS 1212
Db	1162	LRTRFRLEGSRSYQQQPAVSVSIEREQFEEGLI-----ERIQAVVQFPDLE 1210
QY	1213	SEAGWRATLLRGEDDHILTIYMHHTIISDGWSIDVLRLDNLQLYSAALDKSDPLSALTP 1272
Db	1211	RGPLLRVNLOLAEDDHVLIVVOHHIIVSDGWSQMVNVEELVQLY--AAYSQGLDV--LPA 1267
QY	1273	LPIQYSDFAKWKD--QFIEQEKQLNWKYKQLKDDSPA--KIPDTDFARPALLSGDACGVHV 1329
Db	1268	LPIQYADYALWQRSWMEAGEKERQLAYWTGLLGGEQPVLELFDPRPRPARQSHRGQAQGF 1327
QY	1330	TIDGELYQSLRACFNENHTTSFVLLAAFPAAHYRLTAVEDAVIGTPIPNRNRPELEDI 1389
Db	1328	ELSRLEIVRAALAREGASSFMLLIASFOALLYRYSGQADTRVGVPINRNRVETERLI 1387
QY	1390	GCFTVNTQCMRINIDHHDFTGTLINQVKATTTAAFENEDIPFPRVYSALQPGSRDLSSTPL 1449
Db	1388	GFVNTQVLKADQGMGDELLAQARALBAQAQODLPFQOLVLEAQP--ERNASHNPL 1446

RESULT 2

US-09-815-242-5107

; Sequence 5107, Application US/09815242

Db 2113 EYTDLDQRLGCLTYSRDLFDEPRIARMAGHWQNLLEALLGDPORRIAEPL---FRAEE 2169
QY 1548 KLDVNLKVDYDRESSLAD---VFQTVSAYPDSLAVVDSSCLTYTDLDRSDILAG 1603
Db 2170 RKOLLACTAG---EAGLQDPLHGLFAARVAASQAPALTEFAGOTLSYAEELDARSNRLAR 2226
QY 1604 WLRSSMPAETLVAVFAPRSCETIVAFGVKLANLAYLPLDVRSRQVQDIL---SGLSG 1661
Db 2227 VLRSHGVGPEVRVGLALERSLEMVGLLAILKAGGAYVLPDEPLYERLQYMIEDSGVR- 2285
QY 1662 PTIVLIGHDTAPDIEVTNVEFVRIRDALNDSNA-----DGFEVIEHDSKPSATS 1712
Db 2286 ---LLLSH-----AALFEALGELPAGVARWCLEEDGPDALDAEDPAPLAALS 2328
QY 1713 ---LAVLYTSGTGPKGKWMIEHRVIR---TVTSGICPNYPSETRMAMATIAFDGAS 1766
Db 2329 GPQHQAYLIYTGSGTGPKGVAVSHGEIAMHCAAVIECF-GMRAEDCDELHFYSINFDAAS 2387
QY 1767 YEIYSALLFGRTL---CVDYMTLLDARALKDVFFREHVN---ASHVTSSQDVP 1816
Db 2388 ERLIAPLLCGARVVRAGGOWGABEICEILRAEGVSILGTPTPSYGSQAOWLESQGRPLP 2447
QY 1817 LRVPRRLSRTLMFPLVVTDTSTAPDALDAQLYQG---VOCYNGYGPTEGVM---STIYP 1871
Db 2448 VR-----MCITGSEALTGEHLQRIQAFAPAFASFFNAYGTETVVMPLACLAP. 2494
QY 1872 IDSTESFINGVPIGRALNNGSAYVVDPEOOLVGTVNGELVVTGDLGARYSDK-ALDEN 1930
Db 2495 -ERLEEGAASVPIGSVVGARVAYTLDADLALVPOGATGELYGVGAGLARGYHERPALSAE 2553
QY 1931 RFVHITVNDQTVKAYRGDRVYRIGDGLIEFFGRMDTQKIRGNRESAEIEAALRDS 1990
Db 2554 RFPDPPFAEGGRLYRTGDLVRL-CDNGQVEYVGRIDHQVIRGRIELGEIEARLLEHP 2612
QY 1991 SVRDAAYVLAQNEADOAPEILGFV---VADHDHSENDKGQSANQVEGWQDHFESGMYSDIG 2047
Db 2613 QVRE-ALVLAIDSPSGKQLAGYASVAEQD---EDAQA----- 2648
QY 2048 EIDPSTIGSDFKGWTSYDGSQIDFDMHEWLGETTTTLHDNRSLGNVLGTCGSMILF 2107
Db 2649 ---LREALKT----- 2655
QY 2108 NLSRLESYVGLPERSAAAFVNKATESIPSLA-GKAKVOVGATDIDIGVDDLDPLVL 2166
Db 2656 HLKQOLPDY--WPAHLL-----LLASLPLTANGKLDRLALPAPD-----PAL 2696
QY 2167 NSVIQPPSSEYLAETLHLPLNVQRIFFGD-----VRSQATNEHF- 2209
Db 2697 NRQAYEAPRVLQQLAGVWREVLNVERVGLGDNFFELGDSILSIQVVSRAKOLGIFHS 2756
QY 2210 -----LAARAIHTLGNKATKDDYR-----QKMAELEDMEEBELV 2243
Db 2757 PRDLFOHQTVQSLAAVARHSQASQAEQGPVQGSALTPIQHWFFDLPLARHEHWNQALL 2816
QY 2244 EPAFFTSKDRFGLVHEVILPKME-AVNELSAVRYAA-----VVHVRSLGDELV 2295
Db 2817 QPQRAIDL-----GLLKRSQRLVEQHDAURLAFQVDGEWLAQHRPLRQELL 2865
QY 2296 --LPVEK-DWIDFQANQLNOKSL---GDLKSSDAAIMAVSKIPFEITAFERQVVASL 2348
Db 2866 WHVPVQSDFCAELFAKA--QRSILDLEQGPLRA-----VLVDGPGAGEORLLAI 2913
QY 2349 NSNIDE---WQL-----STIRSAGDSSLSVPDIFRIAGEAGFRVVSARQWSONGA 2399
Db 2914 HHLVWDGVSWRVLLEDLQVYRQFAEG-----AEPALPAKTSAFRDWA--GR 2958
QY 2400 LDAVFHCCSQ-----GRTLNVFPDTHHLRGSD-----LLTNRPLO 2435
Db 2959 LQA---YAGESERELGWWQALRGGPVVEWPCDRP-QGDNREALAESVSLRDPQTRQ 3014
QY 2436 RLQNRRIA---IEVRELRSLPSYMI-----PSNIVVLD----- 2467

Db 3015 LLOQAPRAYRTQVNDLLLTALTALRVLCRWSGOPSTLVQLEGHREALFDIDILTRSVGNFT 3074
QY 2468 -----KMLNANGKVRDKELSRRAKVVPK-----QOTAAPLPPTPTISE 2505
Db 3075 SAYPLRUTPAQSPGE-SIKAIKEQLRAVPHKGLGVGLYRLADPAVRQAMALPTAPIT- 3132
QY 2506 VEVILCEEATEVFMKVDITDHFENLGGHSLLATKLISRIDORLKVRTVKVDVDPVFA 2565
Db 3133 -----FNLYLQ-----FDQS-FA 3144
QY 2566 DLASVIRQGLGLOQVSDGQDSAHMAPTETETAILCDEFKAVLGFQVGTITNFFDLG 2625
Db 3145 D-----ALFQPLDOPTG-----PIHDSOAPLPNELS--VDGQV-----YG 3177
QY 2626 GHSLMATKLAVRIGRLDITTVSVKDFDHPVLFOLAIA-----LONLVQSKTNEIVGGREM 2681
Db 3178 GELV-----LRWTVS-RERYDARTVNELAQYLAELQALIEHCLEDGAGGL-- 3222
QY 2682 AYSFPOLLFTEDPEEFMASEI-KPOLELQEI-----TODIYPTSTOMOKAFLEDHTTARPR 2736
Db 3223 -----TPSDFPLAQLSOQALDALAVPAGIEEDVYPLTPMOEGLLL-HTLLEPG 3269
QY 2737 PFV-----PFVIDPSTSEPDAAGLIKACESLVNHLDFRTVEAEAGE-LYQVVLSCLD 2790
Db 3270 TGIYMQDRYRIDSPLDPERFAA-----AQWVARHEALRASFWNAGETMLQVHIKPR 3325
QY 2791 LPIQVETEDINTATNEFLDEFAKEPVRILGHPLIRETIIKQTKSMRVIMRISHALYDGL 2850
Db 3326 TRIEFLDWSLPELPGHEERLQALHKREAG-----FDLLEQPPFHLRLIRLGEARYWFM 3380
QY 2851 SLEHVVRKLHMYNG--RSL-----PPHQSRMYQYTA-----D 2883
Db 3381 MSNH-----HILIDAWCRGLLMDNDFEYISALGESRPAFLTPTP-----RYRDIYAWLQQR 3431
QY 2884 GRESGHFWMDVIO--NTPMTILSD-----DTVVGDNDATCKALHLSKIVNIP 2929
Db 3432 DLQSRWNSLSESLRGERPTLVPSDRFLREHAGESGMVGDRTYTRLDADGARLRELA 3491
QY 2930 SQVLGRSSNIITQATVFNAAACALVLSRESKQVFGRIYVSGRO-GLPVEYQDIDVGPCTN 2988
Db 3492 QRYO-----LTVTFAQAAWALTLLRFSGERDVLFGVTVAGRVPVGM-EMQRTVGLFIN 3544
QY 2989 AVPVRAHIES-----DYNOLLHDIQOQYLLSLPHETIGFSDLKNCNTQWPEATNFSCCI 3044
Db 3545 SIPLRVOMPAGQORCTVREWLNRFLERNLREHEHLPLVAIQES-SELPKQGPLFDSLF 3603
QY 3045 TYHNFYHPSESQEQRVEMGL--TKFVNTEMDE-----PLYDLAIAAGEVEPDGAGL 3095
Db 3604 V-----FENAPVEVSLDRAQSLNASSDSGRTHTNFPLTVVVCYPG-----DDLGL 3648
QY 3096 KVTVIATQTLFGRRKRVREHLLLEESK 3120
Db 3649 HLSY--DQRYFEAPTVERLLGERK 3671

RESULT 3

US-09-976-059-15
; Sequence 15, Application US/09976059
; Patent No. US20020164747A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin
; FILE REFERENCE: 3019-PCT
; CURRENT APPLICATION NUMBER: US/09/976,059
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 4999
; TYPE: PRT
; ORGANISM: Actinoplanes sp.

us-09-976-059-15

Query Match 10.1%; Score 1632; DB 9; Length 4999;
Best Local Similarity 22.5%; Pred. No. 7.3e-109;
Matches 738; Conservative 417; Mismatches 1123; Indels 1000; Gaps 109;

QY 43 SSRIEAIKPCPPFQDMDICNAL-DKQSAIGHAVYDVPTDID-----ISRFALAWKEIV 95
DB 1104 ANIADVYPLAPLOEQIFHHMADRDSA--DVTVTPTVEFSDRDLGDGLAALQQVV 1160
QY 96 NOTPALRAFAFTSDSGKTSOVLTKDSFVSNMCMWSSSSSDPDEVRDEAAAAAGRCNRF 155
DB 1161 DRT-----DVRTSVV-----WGLREPVQVWRHARLPV-----DEV 1193
QY 156 VLLEDM-----QTKKQOLWMTESHVALVDTFOQVLSRVSFAAYKHEKDHRTPESSDAT 211
DB 1194 VLRLDLDPEQLNALGTAW-----MDLS-EAPLVQAVAA-----RPGDPQ----- 1233
QY 212 DTDOSQSVSVMSCEDNAVSAATHFWQTH-----LNDLNASVFPHLSDHLVAPNPTTAE 265
DB 1234 ----EMLAVLRH-----HLVQDHTALDILLEELAAAYLAGRGD---LPEVPVPRE 1277
QY 266 HRIPTPLQKALSNAICTALSILLSRYTHSDEALFGAVTEOSLPE----- 312
DB 1278 F-----VAHRLGVPREEHRYFAGLLGDVTTETAPYGLLDVHSGGLASA 1322
QY 313 DKHYLADGTY--QTVAPLRVHQS-----NLRASDVMDAISYDDRLGLHPLAPFLGRDIRN 365
DB 1323 QAHLRLDGLGRVAFAAREHGVSPATLFLHAWARVLGTLAGRDDVVFCTVLFGRM--- 1378
QY 366 TGDNGSACDFQTVLLVTD-----GSHVNGINGFTLOQTIESHEMPCNNRALLHQC 418
DB 1379 ---NSGAGADRVPGFLINTPLVRVRLGAPVGDALDGLRDLTE-----LIAH-- 1422
QY 419 MESSGALLVAYDHNVID-----SLQTTLLQ----- 445
DB 1423 ---EHAPLAVAQAANLFRGRPLETSIFNYRYARGAEPAGAALDGIKLLSARDLTNYPLAVA 1480
QY 446 -----QFGLIKLQSPDLSS-----MAEVLNLTXYDRABI 477
DB 1481 VDAEGDTFSLTVDAVADPQVQVGGELLYTALRNLTAEANPGTPLAAGVGLGEDELSRV 1540
QY 478 ES-WNSQPLEVQDTHLHEMLKAVSHSPKTAIQAWGDGWTYSELNDVSSRLAVHIKSLG 536
DB 1541 VSGWNDTARRVRQASVPELFAERVAAPGAPAAAGDLRWTYADLDARSALARSILVAG 1600
QY 537 LRAQQAIIIPVYFEKSKWIASMLAVLKSGNAFTLIDPNDPPARTAQVVTQTRATVALTSK 596
DB 1601 VTAEPPVV-VALERSADVLTAFALAVAKAGGVFPVVDLSMPRARVDVIAADCAWIAVADR 1659
QY 597 LHRETQVQLVGRVVDDELQSVASDDFSLSKTSQ---DLAYVIFTSGTDPKGIMI 653
DB 1660 -----PMTGLTVVPANR-----AGDPAVALPPRPLPGAARYMTSGTGRPKGVVT 1706
QY 654 EHR-----AFSSCALKFGASLGINSSTRALQFTHAFGACLEIMTTLINGCVCVI-PSD 707
DB 1707 THQNVVDLTVDRC---WGPT-----PRVLEHAFHAFDASSFELWVPLLTGCTGVVAPGE 1757
QY 708 DDRMNSIPSFINRYNNVMNMTAPSYMGTFSS--PEDVPLGLATLVLGEQM--SSSVNAIW-- 762
DB 1758 SDTGTGLROLIRAHLETHVHTAGLLRVLAEADPSCFAGLEVTLTGDDVVPAEAVRVLDA 1817
QY 763 APKQLLLNGYGOSESSICFASNMTSEPN---NMGRV-GAHSWIDPNDINRLVPIGAV 818
DB 1818 NPGVVRQLYGPTE-VTICATOHVREPSVPLPIGRPLDNTRVYILD--GLLOPPVPGVT 1874
QY 819 GELVIESPGIARDYIVPPPEKSPFFTDIPSYPANTFPDGAKLRYTGDRLARYSDGSIV 878
DB 1875 GELYTAGAGVARGYADMPGTTAERFVAD-----PFTAGRLRYTGDRLVRWTEGELV 1926
QY 879 CLGRIDSQVKIRGORVELGAIETHL-----RQOMPDDLTIWEATKRSQSAN 925
DB 1927 FAGRADDQVKIRGYRVEPGEVEBAVLAALPGVSQAQAVIVREDYPGD-----KR----- 1973

QY 926 STSLIAFLIGSSYFGNRPNSDAHILDHDAFKAINIKLEQVLPHRHSIPSYFICMLPRPTAT 985
DB 1974 ---LVAYLVAAB-----ETVEAARAHAEQRLPSYLVPSFAFVQDLALPLTGN 2016
QY 986 GXIDRRRLRIMGKDILDKQTAIQVQAAPAPIPVADTAALKLHSIHWQSLG-----IDP 1039
DB 2017 QKVDRAAL-----PAPLGFEGAGACAPADAREELVGAFAFVLDL 2056
QY 1040 ATVNVGATFFELGGSNITAKMVMNARSVMGLKYSNIYQHPTLAGISAVKVGDPLSYTL 1099
DB 2057 GRVGDDDDFFALCGHSLALALVRLRQGLGVSVRAVDFDARTPAALTRRGDGGADDRPA 2116
QY 1100 IPKSTHEGPEQSYSGRLWFLDQLDVGSLAWLIPIYAVRMGRPVNVNVALRRLAALEQRH 1159
DB 2117 LRAGARPARPLPSYAQRRLWFLAQLEGPSATYNIPVALRLEGDLDRDALTAALROVVARH 2176
QY 1160 ETLRTTFEDQDGVGVQIVHEKLSER--MKVIDLCSGDLDPPEVLEQEQITPPNLSSEAGW 1217
DB 2177 EYLVRTVTVYADGEPWQHILDPARAEPALVPVDVPAGRVE--EAVAEAAAYAFDLAREIPL 2234
QY 1218 RATLRLGDEDDHILITVMHHIISDGWSIDVLRRLNOLYSAALKDSKDPLSALTPLPIQY 1277
DB 2235 RAVLAPAGDGTIVLVLHHIAADGWSMRPLARDLATAY-AARRRQOAPES--ETLPVQY 2291
QY 1278 SDFAKWOKDOFIEQ-----BKQLNMYKKQKLDSSPAK--IPTDFARPALLSGDAGCVH 1328
DB 2292 ADYALWORDLLGSDSDPASLSRQIAHWRERL-DGVPEELDLPADRPRAAASHRGLHLS 2350
QY 1329 VTIDELVQSLRAFNCNEHTTTSFVLLAAAFRAHRYLRTAVEDAVIGTPTIANNRRPLEDI 1388
DB 2351 AEIPADVHRSRLRAADHGATVMTLQAAAVAVALLSRLGAGTDVPIGTVVAGRADALENL 2410
QY 1389 ICGFVNTQCMRNIDHHDFTGLINOVKATTTAAAFENEDIPFERVVSALQPSRDLSSSTP 1448
DB 2411 VGFVNTLVLRIDTLGDPRDLTVLQGVRELTLALAHQDVPEKLYEELTP-ARSLARHP 2469
QY 1449 LAQLIFAVHSQKDLGRFKFGLESVPVPSKAY-TREDFME-----PHLQETDSLKGSVN 1501
DB 2470 LQVQVMTLDGGGPDGA-ELPGGLAMSVPVTPGAVPAKFDLDTTETFDAAAGEPAGLRVDLI 2528
QY 1502 FADELFKMTENVVVRVFEILLRNGLOSSRTPVSIPLTDGIVTLEKLDVNLVKKHVDYPR 1561
DB 2529 AAADLFDAGTAARLAGYLSRVL--GVLA--DPRRLAEVDPLEAEESRLMAAGEPAPA 2585
QY 1562 --ESSLADVFQTVQSAYPDLSAVVSSCRLLTVELDRQSDILAGWLRSSMPAETLVAVFE 1619
DB 2586 LPEITVAALVAEQCARTPGAVATGPDASLTVAELDERAARTARWLRRHGAGGGAACVCL 2645
QY 1620 APRSCETIVAFFGVILKANLAYLPDLVRSPSARVQDILSGLSGPTIYVLIGHDTAPPDIEVT 1679
DB 2646 MERSAELVAVLLGVMRAGAAYVPDPAIPAERIRFVVTDARAACVV---SESASAGLVDP 2702
QY 1680 NYEVRIRDALNDSNADGFEVIE-----HDSTKPSATSLAYLYVTSSTGRKPGVMIEHR 1734
DB 2703 GYPCCLAI-----DDPAAAAEPAEPDGDGDAAGRPDDPAYIIYTSSTGTGPKGVVSHR 2757
QY 1735 VIIRTVT-SGCIPNVPSETRMAHMTIAFDGASYEISALLFCRTLVCDYVMTLTDARAL 1793
DB 2758 NVVALLTATRLPFGFAGDEWWSHFSAVDFSWELWLGALTGHGRRVVVVVPIAVSRSPDF 2817
QY 1794 KDVFREHNAASHVTSQSDQVPLRV---PRLSRLTMMFFELVVTDPSTADPADLAQGLYQ 1850
DB 2818 WELVVRREGVTVLSQTPSAFAQLMAAADDDRDALRFVVF-----GGEALD-PGLA 2867
QY 1851 G-----VQCYNGYGPTEGVNGVMSTIYPIDSTESFINGVPTGRALNNGSAYVVDPEOOL 1902
DB 2868 GMLARRPDKPRLVNMVYGITETTVHTTYQHIAPGTT---GSVIGRGLPGFGLYVLDALRP 2924
QY 1903 VGIYVMEGLVVTGDLGARGYSK-ALDENRFVHTIYVNDQTVKAYRTGDRVRYRIGDLTE 1961
DB 2925 VPAGVPEGYARQPQVARGYIGRPLTAERFV-ASPFAPGERMYRTGDVARM-TADGRVL 2982

```

QY 1962 FFGRMDTQFKIRGNRIESAETEAALLRDSSVDAVAVLQONEDQAPETILGFVADHDS 2021
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2983 FAGRSDDIKIRGRIEPEVEAVLAAAGPGVQAIVIRE----- 3032
QY 2022 NDKGOSANQVGGWQDFHESGMYSDIGEIDPSTIGSDFKGMTSMYDGSQIDFDEHWEHLGE 2081
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3023 -----DVP----- 3026
QY 2082 TTRTLHDNRSLGNVLEITGSGMILFNLDLSLESYVGLPEPSRAAAAFVNKATESIPSLAG 2141
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3027 -----DKRLVAV----- 3034
QY 2142 KAKVQVGTATDIGOVDDLHPLVVLNSVIOVFPSSEYLAETADLIHLPNVQRIFFGDVR 2201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3035 -----VG-----GDA- 3039
QY 2202 SOATNEHFLAARAIHTLGKNATKDDVRQKMAELEDMEELVEPAFTSLKDRPGLVEH 2261
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3040 ----- 3039
QY 2262 VEILPKNEAVNELSAYRAAVHVHVSGLGDELVPVEKDDWIDFQANQLNOKSLGDLK 2321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3040 ----- 3039
QY 2322 SSDAAIMAVSKIPPEITAFERQVVASLNSIDEMQLSTIRSSAEGDSSLVDPDIFRIAGE 2381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3040 -----ETLRSQA----- 3047
QY 2382 AGFRVESSARQWSQALDAVAFHCCSQGRFLVNFPTDHLRGSLLTNRLQRLQNR 2441
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3048 -----QR----- 3049
QY 2442 IAIEVRERLSLPSYMPISNVILDKMPLNANKVDKELSRKAKVVPKQOATAAPLTF 2501
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3050 -----LPGYLVPAFAVELDRPLTVNGKLDNR-----ALPVDY 3083
QY 2502 PISEVEILCEEATEVEGCMKVDITDHFNLGGHSLATKLISRIDQRKVRITVKDFDH 2561
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3084 ----- 3083
QY 2562 PVFADLASVIRQGLGLOQPVSDGQDRSAHMAPRTEAILCDEFKVLGFO-VGTDN 2620
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3084 -----GRDAGGRAPANAREVILCFRAFAEVLGVVERGVEDD 3119
QY 2621 FFOLGGHSLMATKLAVRIGHRLDTTSVKDQVHPVLFOLAIADNLVQSKTNEIVGGR 2680
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3120 FFALGHSLLVSLVERL-RQGISVPVRAFLFTTPPAGLA-----EAVG--D 3164
QY 2681 MAEYSPQLFTEDPEEFMASEIKPOL-----ELQEI-----IQDIYPTQ 2721
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3165 GAVVVPNLI-----PED--AAELTPEMLPLADLTADLAVVASVPGGAANIADVPLAP 3218
QY 2722 MQKAFLEDHTTA-RPRFV---PFYIDPSTSPDAAAGLAKESLVNHLIDFRT-VFAE 2776
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3219 LQGIFFHHMAADRSADVTVTVVEFDSRDLG--GFALAQVQVDRDVTVTSVMQ 3276
QY 2777 ASGELQVVLSCDLPIQVETEDNTATNEFLDEFKPEVRLGH-PLIRFTIIKQTS 2835
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3277 GLREPQVVRHARLPIDEVELHEGTDPA--EOLIALATERVDLDRAPLIRTTAAVPGS 3334
QY 2836 MR--VIMRSHALYDGLSLEHVVRKHLMLYNGR--SLLPPHQSRYMQVYADG--RESGH 2889
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3335 GRWALLRIHVLVODHTLVDVLLGELRAFLFEGRGDELPEVPVREFVAQARLGVPREHE 3394
QY 2890 GFREDVIQNT-TPMTILSDTVDVGNATCKALHLSKIV--NIPSOVLRSSNI-ITQATV 2945
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3395 RYFAELGVDVETAPYIGTEVHGDSGA--AVHSREVDDDLAARHLRLARSIGVSPAAL 3452
QY 2946 FNAACALVLSRESDSKDWVGRVTSVGRQGLPVEYQDVGCTNAVPVRAHIESDYNQLL 3005
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3453 FHLARVLCAGSRDDVVGTVLFCRMMGGAADRVQGLFINTLPRVFLAAGSTRDAL 3512
QY 3006 HDIQDY--LLSLPHETIGFSDLKRNCTDWEAITNFSCCITVHNFEYHPESQEOORVE 3063

```

```

Db 3513 TGLRDQLAGLLVHCHAPLALAQRAAGITDGSPLFASI-----FNVRHNDQDPAASAGLE 3566
QY 3064 MGVLTKFVNIEMDEPLYDLAIAGEVEPDGAGLKVTVIA 3101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3567 -GIRTVISAETNPL-DASI--DVTGDRFAITVNAVA 3600

```

RESULT 4

```

US-09-815-242-12265
; Sequence 12265, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12265
; LENGTH: 2397
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12265

```

```

Query Match      8.7% Score 1406; DB 10; Length 2397;
Best Local Similarity 24.2%; Pred. No. 6.8e-93;
Matches 571; Conservative 394; Mismatches 903; Indels 495; Gaps 94;

QY 212 DTSQSVSVSMSCEDNAVSAHFQWTHLNDLNASVFPHLSDHLMVNPPTTTAEHRTTFP 271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 176 DKNQDASHI-----ALDSNVERLEKNSDI-----HDSYLPKHPFEQALYQ-TYL 221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 272 LSKALSNSAICTALISILSRVTHSD--EALFCAVTEQSLPDKHYLDGTVQTVPALR 329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 222 IDQ--MTSIDMASLAVSVYLANHIMSQQHVDVTLGIHVPSHLPNDLH-----GNIVPLT 272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 330 VHCQSNLRASDNDVATSSYDDRLGLHAPFLGRDIRNTGNGSACDFQTVLL--VTDGSH 387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 273 L-----TIDAKVQCRFT-----DFNKCVLQNMNSLOQC 301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 388 VNNGINGFIQQITESSH-FMPCNNRAL-----LLHCQMESSGALLVAYVDHNV----- 434
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 302 AKSSLS--LETIFHCYHQMPCNDVEDVHQIHAHTSLADTEIFPHQHGFKLIYNSAA 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 435 --IDSLOTTRLLQOFGHLKLCQSPDLSSMAEVNLMTEYDRAEIESNSQPLEVOD--- 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 YDLLSTIETSLVVRNLYLQITTEENGKRTVTDELNLMTERD---IQLYDDINLSLPDLDD 416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 490 -----TLIHHEMLKAVSHSPTKTAIQWGDW--TYSELDNVSSRLAVHIKSLGRLAQQA 543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


Db 417 AOTVVTLFEQO---VEATPNHVAQ-FDGVFIYQTLNARANDLAHR-----LRNOYGV 466
 Qy 544 IP-----YFEKSKWVIAVLKSGNAFTLIDNDPPARTQAQVVTOTRATVALTSKLH 598
 Db 467 EPNDRAVIAEKSEIMIIAMIGVLKAGAYVDPNYPSPDROEYILKDATPKVVIYQAL 526
 Qy 599 RETVOKLVRCVVDDDELQSVAS--DDFSLTKSQDLAYVIFTSGTGDPKGMIEHR 656
 Db 527 YENSKONINH---ID---LNKIAWKINIDNSECNTLEDHAYVIYTSGGTGNPKGLIPIHR 580
 Qy 657 AFSSCALFEGASLGINSOTRALQFTHAFGACGLLEIMTTLINGCCVCPISDDDRWN--SI 714
 Db 581 GIVRLVHO--NHVYPLNEETILLSGTIAFDAATFEIYGALLNGGLIVAKKEQLLNPIAL 639
 Qy 715 PSFINRYNNMMATPSYMGTFSPEDVPGLATL--VLVGEOMSSSVNAIWAPEKL----- 766
 Db 640 EOLINENDVNTMWLTSSLFNQIASERIEVLVPLKYLILIGEV---LNKAWDILLNORPKH 696
 Qy 767 -OLLNGYGOSESSSICFASNKSTE--PNN--MGRV--GAHSWVIDPNDINRLVPIGAVGEL 821
 Db 697 PQIINGYPTENTTTTNYINIPKNVPIPIGKPIGLTHVIMQG---ERRCGVGIPGEL 753
 Qy 822 VIESPGIARDXIVPPPEKSPFTDIPSWYPANTFPDGAKLYRTGLARYASDGSIVCLG 881
 Db 754 CTSGFGLAAGYLNQPELADPKD-----SNI---NOLMYRSGDIVRLLPDGNIDYLY 804
 Qy 882 RIDSOVKIRGORVELGALETHLRQOQPDLLTIVWEATRKSQANSTSLIA-----FLI 934
 Db 805 RKDAQVKINGFIEELSEV-----HALERIOQINKAVVIVQNHQDOQYIV 849
 Qy 935 GSSYFNGNRPDSAHILDHATKAINIKLEQVLPKPSIPSYICOMLEPRTATKIDRRRLR 994
 Db 850 --AYY---EAMHTLSHNKIKS---QLRMTLPEYMPVNFHIEQIPIITINGKLDKALP 900
 Qy 995 IMGKDILDKOTGAIVQAPAPIPVFAADTAALKHSIWNQSLCIDPATVNVGATFPELGN 1054
 Db 901 IM--DYVD-----TDAYVAPSTDEHLLCQIFADILHVN--QVGIHNFPELGGH 946
 Qy 1055 SITALKMNV--MARSGMDLKVSNYQHPHTLAGIS--AVYKGDPLSYTLIPKSPHEGPVEOS 1112
 Db 947 SILKATLVNRIEASTGKRLQIGDGLQKPTVFEQAQIAKVOEQNYEVIPEAIVKDDYLS 1006
 Qy 1113 YSQGLWFLDQDVGSLWLPIDYAVMRGPNVNDALRALAALQORHETLRTFF--EDQDG 1171
 Db 1007 SAQKRMYLKWSNHRKDTVYNVFLWRLSSELNVAOLRAOAVHLIARHEILRTQYIVVWDE 1066
 Qy 1172 VGOVIVHEKLSKEMKVIDLGSGLDPFVFLNOEQT-----TPFNLSSEAGWRATL 1221
 Db 1067 VQRIVADVAD-----FEVNTHTDEQIEIMRQFVAPFNLEKPSQIRVRY 1112
 Qy 1222 LRLGEDDHILTIVMHHIISDGSWIDVLRRLDNOLYSAALKSKDPLSALTPLPIQYSDFA 1281
 Db 1113 IRSPLHAY-LFIDTHIINDGMSNIQMLNDNALYQHL-----LLPLKLOYKDY 1162
 Qy 1282 KWKQDOFTEQEKOLNWKQLKDDSP-AKIPDFARPALLSDGACCVHVTIDGELYQSRL 1340
 Db 1163 EKMSHRDKTKHQ--YWLISOFEKDEVPILSLPTDYVRPNKTTNGAMMSFTMNQOOROLLO 1220
 Qy 1341 APCNEHNTTSFVLLAAFAAAHRLTAVEDAVIGTPIANRNRPEDIEDIGCFVNTQCMRI 1400
 Db 1221 KYVEKHQITDFEFMSVWMLLSRYARKDDVVVGVSVMARMHKGAEQMLGMAFANTLVYRG 1280
 Qy 1401 NIDHHDHDTGTLINQVAKTTAATFENEDIPFERVVSALQPGSRDLSTPLAQLIFAVHSOK 1460
 Db 1281 QPSDPKMWTFQLOEVKENSLEAYEHOEYFPECLVNDLQ--SHDASRNPLFDVWMLVQNE 1339
 Qy 1461 ----DLGRFKFOGLESVPVPSKAYTRFDMFHLFOETSLKGSVNFADLFKMEIVENVV 1516
 Db 1340 TNHAHFGHSHKTHIQ----PKSVATAKFDLSFIEEDRDDYITINIEYTDLYXSTVRHMG 1395
 Qy 1517 RVFFELIRNGLOSSRTPVSILPLTDGIVTLEKLDLVN-----KHVDYPRESSLADVFQTO 1572

Db 1396 NOCM-IMIDYILKHQDITLQICDPNG--TEELNNWNTHVNDMLNPNVGNKSIISYFNEV 1452
 Qy 1573 V5AYPDSLAVDVSSCRLTVTELDROSDILAGLRRRSMAPEATLVAVFAPRSCETIVAPFG 1632
 Db 1453 VSRQGNHVALVNDLITMYETLURNVYDAIAHLLSNGVNGCORVALFTESEFEMTAAMLA 1512
 Qy 1633 VLKANLAYLPLDVRSPSARVQDILSGLSGPTIVLIG--HDTAPPOIEVTNVE-FVIR-- 1687
 Db 1513 TVKVGCASYIPIDIDFPKPKROGAILEDKAVTAVMSYGIETETTLPIVIOLENAKGFVESKEN 1572
 Qy 1688 ---DALNDSNADGFVEIHDSTKPSATSLAYVLYTSGSGTRGPKGVMEIHRVIRVTSOC 1744
 Db 1573 EYDLDHGDQLENTAMLNDE-----MYAIYTSGTGMPKGVAIQORNLNLVHA-- 1621
 Qy 1745 IPNYPSEYRMA-----HMATIAFGASYEISALFGRTLVCVDVMTTLDARALKDVF 1798
 Db 1622 ---WSTELQDGNVEFLOHANIVFDASVMEIYCCLLNGHTLVIPOREERNVPAOLOLIN 1678
 Qy 1799 REHVNAASHVTSQDQVPLVRPRRLSRTLMPFEL--VWTDSTAPDALDAQGLYQVQC-- 1854
 Db 1679 KHRVTVAS-----IPLCMCVMEDEFYIEKLITGGATSTASFVK--YIEKHCGT 1724
 Qy 1855 -VNGYGPTEGVMSTIYPIDSTESFINGVPIGRALNNSGAYVVDPEQQLVGIVGMCELV 1913
 Db 1725 YFNAYGPSESTVITSWSHQCGLDIPETIPGRPLSNIOVIYIMS-EGLLCGIGMPGELCI 1783
 Qy 1914 TGDGLARGY-----SDKALDENRFVHITVNDQTVKAYRTGDRVYRIEDGLIEFPGRM 1966
 Db 1784 AGDSLAIQVYNRPMLMADK-WQNPF-----GKGLYHSGDLARY-TSDGOIEPLGRI 1834
 Qy 1967 DTQFKIRGNRIESAIEAALLRDSSVYRDAAVVLOQNEOQAPILGFPVADHDSHENDKQ 2026
 Db 1835 DKQKVNQYRIEIDETENAILAIRGISDCVTVVSHFDTH--DILNAYVYGEQVEODLQ 1892
 Qy 2027 SANQVEGMDHPESGMYSDIGEIDPSTIGSDFKGWTSMYDGOIDEDHEWELGETRTL 2086
 Db 1893 YLN-----DHLPKYMI-----PKTI-----THIDCMLP-----JTNDK 1920
 Qy 2087 HDNRSLGNVLEIGTSGMILNLDLSRLSYVGLSPSRAAFVKNATESIPSISLAKAKVQ 2146
 Db 1921 VDTTRLNPFSP-----QOSKNVYSEPS-----NEIEQTFVDVFGVQLQ 1960
 Qy 2147 VGTATDQGVDD-----LHPDLVY-----LNSVIQY-----PPSSE 2177
 Db 1961 ----NDVGVDHDFELGGSNLEAMLVYSDLKRFHHISMOTLYQYKTVQIVNYMHQNOQ 2016
 Qy 2178 YLAETADTLHLPNV--ORIFFGCDVRSOATNE-----HFLAARAIHTLGKNATK 2224
 Db 2017 SLVALPDNLSELQNVWTRYNLGILEDSLSHRPLGNTLLTGATGFLGAYLIEAL----- 2070
 Qy 2225 DDVRQKMAELEDMEBELLVEPAFFTSKDRFPGLVEHVEITLPKNMEAVNLSAYRYAAV 2284
 Db 2071 QGYSHRIYCFVRADNEEIAWYKILMTNLNDYFS--EETVEMLSNIEVI----- 2116
 Qy 2285 HVRSGLG-----DELVLPEKDDWI-----DFOANQLNQKSLGDLKSSD---A 2325
 Db 2117 ----VGDPECDDVVVLPENMDTIHAGARTDHFQDDDEFEKVNVOGTVDVIRLAQOHA 2171
 Qy 2326 AIMAYSKI-----PFEITAF--ERQVVASLNSNID-- 2353
 Db 2172 RLIIYSTISGVYFDIDTETDVTFSADYVKGQLLISPYTRSFYSELKVLAEVNNGLDGR 2231
 Qy 2354 -----EMQLSTIRSS 2363
 Db 2232 IVRVGNLTSPYNGRWHMRNKTN 2254

RESULT 5
 US-09-976-059-14
 ; Sequence 14, Application US/09976059
 ; Patent No. US20020164741
 ; GENERAL INFORMATION:
 ; APPLICANT: Farnet, Chris


```

; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin
; FILE REFERENCE: 3019-PCT
; CURRENT APPLICATION NUMBER: US/09/976.059
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 4999
; TYPE: PRT
; ORGANISM: Actinoplanes sp.
; US-09-976-059-14

Query Match      8.5%; Score 1375.5; DB 9; Length 4999;
Best Local Similarity 28.0%; Pred. No. 4e-90;
Matches 473; Conservative 267; Mismatches 690; Indels 257; Gaps 55;

QY 277 LSNSAICRTALSILLSRYTHSDEALFGAV-----TEQSLPFDKHYLADGYQTVAP 327
DB 3471 VSPATVPHLAWARVLSVIAGRDVVVFGTILFGRMNSGAAAEVVP-----GLFINTLP 3522
QY 328 LRVHCQSNLRASDYMDAISYYDDRLGHL-----APFGLRDIRNTGDMGSAACDFOTVLLV 382
DB 3523 VRV-----RLNGTSVGEALTALRDQMAELMAHEHAPLALAOAGGVPAGS-----PLFTSLFN 3575
QY 383 TDGSHVNGINGFLOQITESHFMPCNNRALLLHCMESSGALL-----VAYYDHNVID 436
DB 3576 YRNVAGGGDGGGALLEGVTPVLRHTTNNPVVV-----SVDDDDTSFDLVVEAVAPAEAGRVG 3632
QY 437 SLOTRLLQFGLHKLKQSLPDLSSMAEYNLMTEYDRAETE-----SWNSOPLEVQDTLI 492
DB 3633 RLMECELAELVGLAGAPETPL-----SRVRVI-----DEAEIERVYHWNDDTARPVVSSV 3684
QY 493 HHEMLKAVSHSPKTKTAQWGDWGTYSSELDNVSRLAVHIKSLGLRAQQAIIPIVFEKSK 552
DB 3685 PALFAEQVAAPADATAVGEGVSWSYRELDARSALRSVAAGVGESPVV-VALERSP 3743
QY 553 WVTASMLAVLKSGNAFTLIDNPDPARTAOVTTQTRATVALTKSLHRETIVOKLGRGVVV 612
DB 3744 EVLSAFLAVAKAGGVFPVLDLWSPQARIDAVADCAARAVADR-----PMSGLTWP 3796
QY 613 DDELLQSVASDDFSSITKQDGL-----AYVIFSGSGDPKGMIEHRAFSSCALKFGAS 668
DB 3797 ADQV-----GSAVVLPAQVPGAAVRYMTYSGSGRKPQGVVTHQNLVLDLA----- 3843
QY 669 LGINSDT-----RALQFGTHAFGACILLEIMTTLINGGCVCV-PSDDRMNSIPSFNRY 721
DB 3844 ----TDFCWGPTPRVLFPHAFADASSYEIWPVLLNGTGVVAPORSIDATVLRDLIRGH 3899
QY 722 NVNWMATPSYMGTFSPEDVPGLATLVLGEQWSS-SVNAIWA--PKQLLLNGYGQSESS 778
DB 3900 ELTHVHVHTAGLLRLVLDPSCFAGLTEVLTGGDAVSAEAVRRVREANPGLRVQLYGPTE-V 3958
QY 779 SICFASNMSTENPMNGRAV-GAHSWVTDNDINRLVPIGAVGELVIESPGIARDYIVPPP 837
DB 3959 TLCAQHLLVDGVPVIGRPLDNTRVYVLD--DLQPPVPVGTGELYVAGAGLARGYAGMPG 4016
QY 838 PEKSPFETDIPSWYPANTFPDGAKYRTGDLARYASDGSIVCLGRIDSQVKIRGORVELG 897
DB 4017 LTAERFVAD-----PFSVGGRLYRTGDLVRMTDDGVLHFAGRADDDQVKIRGYRVEPG 4068
QY 898 AIETHLRQOQMPD--DLTIVVEATKRSOSANSTSLIAFLIGSSYFGNRPNSDAHILDHATK 955
DB 4069 EVEAVLAQH--PDYSQVAVVV-----REDTPGDKRLVAVVVGDD----- 4105
QY 956 AINIKLEOVLPRHSIFSYICMLELPTATGKTDRRLRMGRDKILDKQTOGAIVQOAPA 1015
DB 4106 -VEAYAOERLPGVLFPSAFVHLDALPLTNSQKQVDRAL-----PA 4144
QY 1016 P-----IPVFADTAALKIH-SIWVQSLGIDPATVNVYGATFFELGNGSITAIKMYNMAR 1066
DB 4145 PSVESGVGRAP--ADAREELMCAAFAEVLDLD--RVGVDDDFALGSHSLVLYLVGRIR 4200

```

```

QY 1067 SV-GMDLKVSNIYQHPTLAGISAVVKGDPGLSYTLIPKSTHEGVPBQSYSGRLWFLDQID 1125
DB 4201 QVFGVEVSARLVDFDARTPAGVVVARLSEGTAREAVRVRPARVPLSLFAQRRLWFLSQLE 4260
QY 1126 VGSWLWYLIPIYAVRMGRPVNVVDALRRALAAQSRIETLRTTFEDQDQGVQVIVHEKLSSEM 1185
DB 4261 GPSATYNIPVALRDLGDLDRDALTAAALHDVVVARHEVLRVFTVADGEPWQQLDDPPQSV 4320
QY 1186 KVIDLGSDDLDPFEVLNQEQTTPENLSSEAGWRATLLRLGEDDHLTLTMHHILSDGWSI 1245
DB 4321 PVVEVTFDRLP--FAVAAACHRFDLGRELPRLAVLLATGDDVHVLLVHVHIAADGWSM 4378
QY 1246 DVLRRDLNQLYSAALKOSKDPLSALTPLPIQYSYFAKQKQDFTEQ-----EKQLNYW 1298
DB 4379 RPLARDLAAVAARI-DATAP--ALGALPVQADYALWQDVLGSEHPDPSVISQOVAYW 4435
QY 1299 KQKQKD--SSPAKIPTDFARPALLSGDAGCVHVTIDGELYQSLRAFPCNHNHTSFVLLAA 1357
DB 4436 RRLAGVPEELDLDPVDRARPAEASHRGHTVEFAVPPVHHQALAEARRNGVTVMVTQTA 4495
QY 1358 FRAAHYRLTAVEDAVIGTPIANRNRPEDLIIIGCFVNTQCMRINIDHHDTEGTLINQKA 1417
DB 4496 LAVLLSKLGACTDPIGVAVAGRTDPTLDNLIGFFVNTLVLTDLTGNTPTITDLLHTRD 4555
QY 1418 TTTAAFENEDIPFERYVSALQPSGRDLSSTPLAQLIFAVHS-QKDLGRFKFGLESVPVP 1476
DB 4556 TTLHAFTHQDVPFEKLVEDLAP--TRSLARHPLFQVMMTLOSASADEEPLALAGLRVTDLP 4614
QY 1477 ---SKATYRDMERHFLQETDLSKGSVNFADLKFMTETVNVVRVFEILRNGLOSRTP 1533
DB 4615 AGETPAKVLDLTLHEVAGRDMHATLLGAADLFEQETVRALAD---RLRTLEMAAAP 4671
QY 1534 VSILPLTDGIVTLEKLDVFN-----VKHVDYPR---ESSLADVFOFQVSAVPSDSLAV 1582
DB 4672 DB-----RLDRIEVLSPGERSLLVENNDTARPVVSSVPALFAEQVAAAPDAVAV 4722
QY 1583 VDSSCRITYTELDRQSDILAGWLRRRSMPAETLVAVFAPRSCETIVAFFGVKLANLAYLP 1642
DB 4723 VGEVSWTYRELDARSALARSVAAGVGESPVVVALERSPEVLSAFLAVAKAGGVFP 4782
QY 1643 LDVRSRQVQDILSG-----LSGPTIVL--IGHDTAPDPDIEVTVNFVRIRDA 1689
DB 4783 VDLSPQARVDVAVVAGCARIADRPMSGLTVVVSAGLGGDSAV-----V 4827
QY 1690 LNDSNADGFVEIHDSTKPSATSLAYLYTSGTGRPKGVMIHRVIRTVTSGCIPNYP 1749
DB 4828 SGDLTADRAVVL---PAGVPVGAAYRYMTYSGTGRPKGVVTHQNLVDLATDTCWGTP 4884
QY 1750 SETRMAMATIAFDGASYEIIYSALLFGRTLVCVDYMTTLDARALKDVFREHVNAASHVT 1809
DB 4885 ---RVLPFHAPHAFDASSYEIWWPLNGGTVV-VAPRPSIDATVLRDL-IGAHELTHVHT 4939
QY 1810 SSSQDVLPRVPRRLSRTL--MFFFLVTVTDSTAPDALDAQLYQ-----GVQCXNGVGT 1861
DB 4940 AG-----LLRVLDPSCFAGLTEVLTGGDAVSAEAVRRVKDANPGLRVRLYQYPT 4988
QY 1862 ENGVMS 1868
DB 4989 EVTLCAT 4995

```

RESULT 6

```

US-09-815-242-5654
; Sequence 5654, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.

```

APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011a
CURRENT FILING DATE: 2001-03-21
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5654
LENGTH: 1668
TYPE: PRN
ORGANISM: Staphylococcus aureus
US-09-815-242-5654

Query Match		8.5%; Score 1374; DB 10; Length 1668;	
Best Local Similarity		27.3%; Pred. No. 7.9e-91;	
Matches 459; Conservative 292; Mismatches 683; Indels 246; Gaps 62;			
QY	461	SSMAEVNLMTEYDRAIESWNSQPLEVOD-----TLIHREMLKAVSHSPTKTAIOAW	512
DB	37	TTVDDELNLTEDR---IQLYDDINSLPLDDAQVWTLFEQO-----VEATENHVAQ-F	88
QY	513	DGDW-TYSELNVSRLAVHIKSLGLRAQQAIIIP-----VYPEKSKWVIAASLAVLKSGN	566
DB	89	DGVFTYQTLNARANDLAHR-----LRNQGVEPNDRVAVIAEKSTEMIAMIIGVLKAGG	143
QY	567	AFTLIDPNPPARTAAVQTQTTRATVALTSKLHRETVOKLVGRVVDVDELLOSVSAS--D	624
DB	144	AVTPIDPNTPSRQEVYILKDATPKVVITYQALYENSKONINH---ID---LNKIANKNID	197
QY	625	DFSSLTKSODLAVYFTSGTDPKGMIEHRAFSKALCFGASLGINSRDLALQFGTHA	684
DB	198	NLSECNLTEDHAYVIYTSCTTGNPKGTLIPHRGIVRLVHQ-NHYVPLNEETILLSGTIA	256
QY	685	FGACILLEMTTLINGCGVCIPISSDDRMN--STPSFINRVNVMATPSYMGTFSPEDVP	742
DB	257	FDAATFEIYGALLNGKLIIVAKKEOLLNPALAEQLINENDVNTMLTSSLFNQIASERIE	316
QY	743	GLATL--VLVGRQMSVSNVNAIPKL-----OLLNGYGOSESSICFASNSTF-PNN	792
DB	317	VLVPLKYLIGGEV---LNKAVDNLNQRKHPQIINGYGPTENFTFTTYNIPKNVPNR	373
QY	793	--MGRV-GAHSWIDPNINRLVPIGAGELVIESPGIARDYIVPPPPKSPFFTDIPS	849
DB	374	IPIGKPIGLTHVIMQG---ERRCGVIGBELCTSGFLAAGYLNQPELTADKFKID---	427
QY	850	WYPANTFPDGAKYRTGLDARYASDGSIVCLGRIDPSQVKIRQORVELGAIEHLRQOMPD	909
DB	428	---SNI---NQLMYRSGDIVRLPLDGNIDYLRKDKQVKIRGFIELSEVE-----	472
QY	910	DLTIIVVEATKRSQSANSTSLIA-----FLIGSSYFGNRPDSAILHDHATKAINIKLE	962
DB	473	-----HALERIQGINKAVVIQNHDDQOYIY--AYY-----EAMHTLSHNKIKS---QLR	517
QY	963	QVLPKHSIPSPYICMLPRTATGKIDRRRLRMCKDILDKOTGAIVQOAPAPIPVFPAD	1022
DB	518	MTLPEYMPVNFVNFHIEQIPITTINGKLDKALPIM--DIYD-----TDAYVAPSTD	565

QY	1023	TAAKLHSIWQSLGIDPATVNVGATTFELGGNSITAIKMN-MARSVGMDLKVSNIYQHP	1081
DB	566	TEHLLCQIFADILHVN--QVGITHDFFELGGHSLKATLVNRIEASTGKRLQIGOLLQKP	623
QY	1082	TLAGIS-AVVKGDPLSYTLIPKSTHEGPEQSYSGRLWFLDQLDVGSLLWLYLPVAVMR	1140
DB	624	TVFELAQAIAKVOEQNEIPEIATVKKDDVLSAOKRMVLLWKSNNHKDFVYNWFLWRLS	683
QY	1141	GPVNVDAFLRAALAELQRHETLRTTF-EDQDGVGVQIVHEKLSSEMKVIDLCSOLDPFE	1199
DB	684	SELNVAQLRQAVOHLIARHEILRTQYIVVDDEVQRIVADVAD-----FE	729
QY	1200	VLNQEQT-----TPFNLSSEAGWRATLLRGEDDHIITVHHHIIISGWSIDVLR	1249
DB	730	EYVTHFTDEQETMRQFVAPFNFLEKPSQIRVYIRSPHAY-LFIDTHHIIINDGMSNIQLM	788
QY	1250	RDNLNOLYSAALKDSKPLSALTPLIOYSDFAKQKQDFIEQEKQLNWKYKOLKSSP-A	1308
DB	789	NOLNLYQHKL-----LPLKLQYKDYSEWMSHRDWTKHO--YWLSQFKDEVPIL	837
QY	1309	KIPTDFARPALLSGDAGCVHVTIDGELYQSLRAFCEHNTTSFVLLAAAFRAAHYRLTAV	1368
DB	838	SLPTDYVRENIKTNGAMSFMTNQTRQLLQKYVEKHQITDFMFMSVVMTLLSRYARK	897
QY	1369	EDAVICTPIANRNRPEDDIIGCFVNTQCMRINIDHHDFTFGLINQKATTTAAFNEDI	1428
DB	898	DDVVGVSVMSARMHKGAEQMLGMFANTLVYRGQSPDKMWTQFLQEVKEMSLAEYBQY	957
QY	1429	PFERVVSALQPSRDLSTPLAQLIFAVHSQK---DLGRFKFGLESVPVPSKAYTRFD	1484
DB	958	PEECLVNDLQ--SHDASRNPLFDVMLVQNNETNHAHFGHSLKTHIQ-----PKSVTAKED	1012
QY	1485	MEFHFLQEPEDSLKGSVNADELFKMETVENVYVFEILNRGLQSSRTVSLPLTDGIV	1544
DB	1013	LSFIEEDRDDYITINLEYNTDLYHSETVRHMGNCM-IMIDYILKHQDLOICDIPNG--	1069
QY	1545	LEKLDVLN-----KHVDYPPRESSLADVFQTVQVAYDSLAVVDSRCRTYTELDRQSDI	1600
DB	1070	TEELLNVWVTHVNDRLNLPVGNKSIISYFNEVYVSRQGNHVALVNDLTTIETLRYNYDA	1129
QY	1601	LAGWLRRRMPAETLIVAFAPRSCETIVAFGLKANLAYLPDVRSPSARVODILSGLS	1660
DB	1130	IAHLLNSGNGQORVALFTERSFEMIAAMLATVKGASYIPIDIDFPKROGAILLEDAK	1189
QY	1661	GPTIIVLIG--HDTAPDPIEVNVE-FVRIR-----DALNDSNADGEFVIEHDSKPSATS	1712
DB	1190	VTAVMSYGIETITLPIQLENAGKGVESKENEQYDDLHGDQLENTAMLDNE-----	1241
QY	1713	LAYVLTSGSTGRPKGVMIHVRVIRTVTSGCIPNYPSETRMA-----HMTATPDGAS	1766
DB	1242	-MYAITSCTTGMKGVAIQRNLLMLVHA-----WSTELQGLDNEVFLQHANIVPDASV	1295
QY	1767	YEYSALLFGRTLVGVVYMTTLDAKALKDVFEFREVHNAASHVTSOSSODVPLRVPRLSRT	1826
DB	1296	MEYCCLLNGHILVLPDREERNPAOQLOQLINKHRVTAS-----IPLQMSV	1343
QY	1827	LMFFFL--VYDSTAPDALDAQLYQGVQO---YNGYGTENGVMSTIYPIDSTESFING	1881
DB	1344	MEDFYIEKILITGATSTASFVK-YIEKHCGYFNAYGPSSEVITSYSHOCCGDLIPET	1401
QY	1882	VPIGRALNNGAYVVDPEQOQVIGVMGELVTVGDGLARGY-----SDKALDENRFVH	1934
DB	1402	IPIGPLSNQIVVINS-EGLLCGIGMPGLCAGDSLAIGYINRPELMADK-WQNNPF--	1457
QY	1935	ITVNDOTVAYRTGDRVYRIGDGLIEFFGRMDTQFKIRGNRIEASIEAALLRDSVSD	1994
DB	1458	-----GKGLYHSGDILARY-TSDGQIEFLGRIDKQVKNYRIELDEIENAILAIRGISD	1511
QY	1995	AAVVLQONEDQAEIIGLVFVVDHSDHSEKDGOSANOVQWQGHFEGSMYSDIGEIDPSI	2054
DB	1512	CVVTVSHFETH--DILNAYYVGEQVQEDLKQYLN-----DHLPKYMI-----PKTI	1556

;; TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin

;; FILE REFERENCE: 3019-PCT
 ;; CURRENT APPLICATION NUMBER: US/09/976,059
 ;; CURRENT FILING DATE: 2001-10-15
 ;; NUMBER OF SEQ ID NOS: 34
 ;; SOFTWARE: PatentIn version 3.0
 ;; SEQ ID NO 13
 ;; LENGTH: 1051
 ;; TYPE: PRT
 ;; ORGANISM: Actinoplanes sp.
 ;; NAME/KEY: misc_feature
 ;; LOCATION: (1)..(1)
 ;; OTHER INFORMATION: V represents a non-standard codon. It is expected that the biosynthesis of ramoplanin will have a formylmethionine residue at this position.
 ;; OTHER INFORMATION: sition
 US-09-976-059-13

Query Match 3.9%; Score 623; DB 9; Length 1051;

Best Local Similarity 21.4%; Pred. No. 2.6e-36;

Matches 328; Conservative 157; Mismatches 477; Indels 568; Gaps 42;

1114 SGRFLFDLDQDVGSLWLYLPYAVRMKGPVNDALRRALAEQRHETLRTTFEDDQGVG 1173

12 AQLAVHQAQLAPHSPVQYGEFEVDGECDDLLVAALQVVMGEASRLRFRVIDGTP 71

1174 VQIVHEKLSSEMVIDLQSGSDLPFEV---LNQEQTPFNLSSEAGWRATLRLGEDDH 1229

72 WQYVAEDGDDPIQVVDL-GAAADPRAAALGRMAADLRDGLRDPGLVEHHVLLGEGRV 130

1230 ILTVHMHISDGSIDVLRDLNQLYSALKSKDPLSALTPLPIQYSDFAKWKDQFI 1289

131 IWHFRAHIVCDGSLGIVASRVAGVYSAAGDGVRRPGALPLPLSLLSAADAYERSG-- 188

1290 EOEKQLNWKKOLKSSPAKIPDFAFPALLSG-----DAGCVHVTIDGELYQ 1337

189 DRDRDRHWRSA-LA-GLPAELLACAGRPPLPGPPVRRHEHDLAAEAG----- 235

1338 SLRACNEHNTTSFVLLAAFAAHYRLTAVEDAVIGTPIANR-NRPELEDIIGCFVNTQ 1396

236 RLKAGARRLTSVAQAGIAAALYQHRLTGARDVLVAVPVAGRTTRPEF-DVPGMTSNV 294

1397 CMRINDHHDFTGLNNOYKATTAFAENEDIFERVMSALQPGSDLSSTPLAQLIFAV 1456

295 PVRLAVTPATTVGELLRDVARGVRDGLRHQRYPNIVD-----DLGLADRAALRPVT 347

1457 HSQKDLGR-FKF-----QGLSEVPVPSKATRYTFDMERHFLFOETDSLKGSVNFADFLK 1508

348 VNALALGRPLRFGSAGVRSGLSAGPVD-----DVTIGLYEKVSG--GGMTIATLNP 398

1509 MET-----VENVVRVFEILRNGLOQSSRT--PVSILPLTDGIVTLEKLDVLRKHVDYPRE 1562

399 GRTRDPDAAEVSRWFTLLR-GLAESDAGDPVARIDIVDEPERRLLDENNTAA--PSS 455

1563 SSLADVPTQVSYAYPSLAVVDSRCLTYTELDRQSDILAGWLRRRSMPAETLVAVFAPR 1622

456 DTVLARFEEQAARTPEAPVVCVDVTVAELEGANRLARVLARGAGPESVVALCLPR 515

1623 SCTIVAFFGVLANLAYLPDVRSPSARVODILSGLSGPTIVLIGHDT--APPDIEVTN 1680

516 GPEVVTGILAAWAGAAAYLPVDTELPARVAYLL-GDSAAAARLGTAEALPAALPDGPAAD 574

1681 VEFVRIRDALNDSNADGFEVIEHDSKPSATSLAYLYTSGTGRPKGVMIHRVIIRT 1740

575 VD-VHAPEIARESPP-----LRLEPLDQLAYLYITSGTSLKSGVGVSH----- 619

1741 TSCICPNPSETMAHMTATFDGASYEIYALLFGRT-----LVCVDYMTLDA--- 1790

620 --GGLANY-----VGMASLYGLGLAPLHSSLAFLDTVTSVFVPLVCGSVGVVSAAGGG 671

1791 RALKDVFREHVAASHVTSQSDVP--LRVPRRLSRTLMFFFLVVTDSAPDALDAQG- 1847

672 RGLASLL-----AAGDGFSLVKVVPQHLLR-----LIALVPAAGMAAVGSLVAGGE 717

QY 1848 -----LYQVQVCYNGYPTENGVMSTIYPIDSTESFINGVPIGRALNNSGAYV 1895
 DB 718 VLAGGDVRELSRVPGSVVNVYGTPTETVVCSCVFSVAAGDVVDVVPVGRVANTFLFV 777
 QY 1896 VDPEQQLVGIGVMGELVVTGDLARGYSKKA-IDENRFV--HITVNDQTVKATYTGDRVR 1952
 DB 778 LDEGLRPVPAGVAGELYVAGSOVARGVYGRSGLTASRFVACPFVGGE---RMYRTGDVVR 834
 QY 1953 YRIGDGLIEFFGRMDTQFKIRGNRIESAETEAALLRDSVRAAVVLQQNQDQAPETLGF 2012
 DB 835 LAGGD--LVFVGRVDEQVKIRYVE-----PDEVRLL 864
 QY 2013 VVADHDHSEKDKQGSANQVEGWQDHFESGMYSDIGEIDPSTIGDFKGMTSMYDGSQIDF 2072
 DB 865 VWAGH----- 869
 QY 2073 DEMHNLGETTRTLHDNRSLGNVLEITGTGSMILFNLDLSRLSESVVGLPESRRAAFVNSKA 2132
 DB 870 ----- 869
 QY 2133 TESIPSLAGKAKVQVGTATDIGQVDDLLHPLVVLNSVIOYFSPSEYLAEIADTLIHLPNV 2192
 DB 870 ---PRVAGAAYV-----ARPDV----- 884
 QY 2193 QRIFPGDVSQATNEHFLLAARAHTLGKNATKDDVRQKMAELEDMEBELLEVPAFTSLK 2252
 DB 885 ----- 884
 QY 2253 DRFPGLVHEVILPKMNEAVNELSATRYAAVHVHVRGSLGDELVLVPEKDDWIDFQANQLN 2312
 DB 885 ----- 884
 QY 2313 QKSLGDLKSSDAAIMAVSKIPEITAFERQVVASLNSNIDENQSLSTRSSAEGDSSLSV 2372
 DB 885 -----GERQLVAYVVA----- 895
 QY 2373 PDIFRIAGE-AGFRVEYSSAROWSONGALDAVFHCCSOGKRTLVNFPTDHLHRSGLLTN 2431
 DB 896 -----AGEPAG----- 901
 QY 2432 RPLQRONRRIATEVRELRSLPSYMIPSNIVVLDKMLNANGKVDKELSRRAKVVPK 2491
 DB 902 -----LAESVRAHVAERLPEYMPVPAVTVLDEIPLTVNGKVDRAALPE-----PG 946
 QY 2492 QQTAAALPTFPISEVEVILCEATEVEFGM-KVDITDHFHNLGHSILATKLISRIDQRLK 2550
 DB 947 PVATGNADREPTTERESLLCGAFADVLCIERVGVDDDFSLGHSLLATSLVSRVRLVIG 1006
 QY 2551 VRITVKDVPDHPVAFADLASVIRQGLGQQP 2580
 DB 1007 EELPIELFATPTPAELAAWLQNRADRPQ 1036

RESULT 9

US-10-166-087-44

; Sequence 44, Application US/10166087

; Publication No. US2003007767A1

; GENERAL INFORMATION:

; APPLICANT: Ecopia Biosciences Inc.

; APPLICANT: Farnet, Chris

; APPLICANT: Staffa, Alfredo

; APPLICANT: Zazopoulos, Emmanuel

; TITLE OF INVENTION: Genes and proteins for the biosynthesis of anthramycin

; FILE REFERENCE: 3014-2US

; CURRENT APPLICATION NUMBER: US/10/166,087

; CURRENT FILING DATE: 2002-06-11

; NUMBER OF SEQ ID NOS: 51

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 44

; LENGTH: 1446

; TYPE: PRT

; ORGANISM: Streptomyces refuineus subspecies thermotolerans

US-10-166-087-44

Query Match	3.88;	Score	612.5;	DB	9;	Length	1446;
Best Local Similarity	25.4#;	Pred.	No.	2.7e-35;			
Matches	293;	Conservative	163;	Mismatches	519;	Indels	177;
Gaps	40;						
QY	1101	PKSTHEGPEVQSYSGRLWFLDQLDVGSLWYLIPYAVVRMGPNVNDALRALAALEORHE	1160				
DB	7	QOATDTRCFPPSPAQAGLWFASYGTDPTAYNOLPLVRLUGTLVDHTLLHRLRLVHREHC	66				
QY	1161	TLRTTFE-DODGVGVQIVHKEKLSBEEMKVIDLCCGSDLPF--EVLNQBOTPTFFNLSSEAGW	1217				
DB	67	ALRTTFDMADGELRQIVHCELEPIVDVVRHAGDSSEAWAEQVEQVAATVEDLRRGPLA	126				
QY	1218	RATLLRL-GEDDHILITVMMHIIISDGSIDVLRDLNQLYSAAKDSKDPLSALTPLPQ	1276				
DB	127	RVRLHLRLVABGRSLLVFNHHTVFDGLSWKPYLSRLAEVYTALARGOEPKPRK---PRRQA	183				
QY	1277	YSDFAKWKQDQIEQEKQLNWKYKQLKDDSPAKIPTDFAPALISGDAGCVHVT-----ID	1332				
DB	184	VEAYARWSEW--ADSGLSLHWLDKLD-APA-----AAPVGLPGSEPARHVTTHKAVLD	233				
QY	1333	GELYQSLRARC-NEHNTTTSFVLLAFARAHYRLTAVEDAVIGTPIANRNRPELIDIGC	1391				
DB	235	DRLSAQVKTCATEGTTGMFFAALAPVLHL-RHTGDDILLGVPTVRSGDAE-VYGH	292				
QY	1392	FVNTQCMRINIDHDTFGTLINOVKATTTAAFNEDIPFFRVYSAL-----QPGSRDLS	1445				
DB	293	LTNTVVLRLHLPAGATARDVLHAVKRDMLDRLRHVPLEAVVGLRALGGKDGVDGLF	352				
QY	1446	STPLAQLIFAVHSQKOLGRKFGOLESP-----VPSKAYTRFDMFHLPEQDTSL	1496				
DB	353	NAMLT-VMPASARLDREWGVEFHWHSVGAKYELAVVDETPGRYTL---VVEHTSAS	408				
QY	1497	KGSVNFADELFKMETVENVVRVFEILRNLGQSSRTPVSLPLTDGIVTLEKLDVLNVKH	1556				
DB	409	AGAGSLAAYLAR--RLETLVR-----SVWADPD-----TDVRLRLW	442				
QY	1557	VDYPRESSL-----ADVFQVQVAYPSDLAVVDSCLTYTTELDRQS	1598				
DB	443	VSAAEERAVTGLCARRODAPELGEVTADLFAEAAAALAAAADPAVAVDGVVTSYAEALARQA	502				
QY	1599	DILAGWLRRSPAEITLVAFAPRSCCTIVAFGLVKLANLAYLPDVRSPSARVQDILSG	1658				
DB	503	DAVADLAARGVRDGRPVAVLMPRGLDLVATVVGILRAGGSYVVLDDADPERLSFALAD	562				
QY	1659	LSGPTIVLIGHDTAPPDIEVTNVEFVRIRDALDINSNADGFE--VIEHDSKPSATSLAYV	1716				
DB	563	-SCAKILL--HD--PDADLAG--VRLPDGMQTATMPGTGGVVLPEGRKSPDDQVYV	613				
QY	1717	LYTSGSTGRPKGVMIEHRVLIIRVTSCIPNYSPETRMAMHATAFDGAIEYIYSALLFG	1776				
DB	614	VYTSGSTGRPKGVVLEPTLTNLVRNOQVLSHRRMETLOYMPPAFDVFVTFLEFGTLCGT	673				
QY	1777	RTLVCVDYMTDLARALKDVFREHVN-----AASHVTSSSODVPLRVPRRLSRTLME	1829				
DB	674	GTLVPPPPHARTDFEALAAALAEQRIERAFVYPVALRELAVALRSSCTRLPDLREVIVTG	733				
QY	1830	FFLVVTDSTAPDALDAQGLYQGVQCYNGYQPTENGVMSTYPTIDSTESFINGVPIGRALN	1889				
DB	734	ERLVVTE---DLREMFRRHPGARLINAYGPSEAHLVSAEWLPADDPDTPAVPPIGRVVA	789				
QY	1890	NSGAYVVDPEQQLVGVGMGELVVTGDGLARGYSDKALDENRFVHIVNDOTVKA-----	1944				
DB	790	GLDARVLLEGDEPAPGVGELCAGVPVSPGY-----ICLPBKTQAMVYDDP	837				
QY	1945	-----YRTDVRVRYRIGDLIEFFGRMDTPQKIRGNRIETSAEIAEALLRSDSVRAAD	1997				
DB	838	FVPQQLMYRTGD--VVVLDPDRLHYRGRADDQIKIRGYRVPPEVEAALERVLHVEAAV	896				
QY	1998	V-LQONEDQAPEILGFVADHDISENDKGSANQVGEHQ---DHFESEMYSDIGEIDP	2051				
DB	897	IAPVAGCHDRA--LHAIEFRVSGEPPSNWRSLGVLPFGMTIPRGITRVDATIPVTPNGKTR	954				

```

Qy 2052 STI-----GSDFKGWNTSMYDGSQIDFDEMHEWLGETTTRTLHDNRSLGNVLEI 2098
      :      ||      |      |      |      |      |      |      |      |      |
Db 955 RALEARLADRAGTEPAGGGMDWT---DCERAIADLWTEVLGHGPATPDD-----DFEL 1006
      :      :      :      :      :      :      :      :      :      :      :
Qy 2099 GTCSGMLFLND---SRLESYGLPEPSRAAFVNKATESIPSLACKAKVOVGTAT-DI 2153
      |      :      |      |      |      |      |      |      |      |      |
Db 1007 G-GHSLLAARLRLVQRRLSDVPLS-----VLLGTPTVRGMAGSLAGRGASGTVDL 1057
      :      :      :      :      :      :      :      :      :      :      :
Qy 2154 GQVDDLHPDLVV 2165
      :      :      |      |      |      |      |      |      |      |
Db 1058 REEARLH-DLVV 1068
      :      :      |      |      |      |      |      |      |      |

RESULT 10
US-09-815-242-10079
; Sequence 10079, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10079
; LENGTH: 1293
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10079

```

Query Match.	3.5%;	Score 559;	DB 10;	Length 1293;
Best Local Similarity	25.6%;	pred. No. 1.8e-31;		
Matches 249;	Conservative 166;	Mismatches 416;	Indels 142;	Gaps 41;
QY	1114	SQGRLWFLDGL-DVGSILWYLPYAVMRGPNVVDLRRALAALEORHETLRTTFEDODGV	1172	
Db	10	AQGIWMAEKLSELPNSAWVAHY-VELTGEVDSPLARAVVAGLAQADTLRMRFTEDNCE	68	
QY	1173	GVOIVIEKLSSEEM-KVIDLGGSDLDFP---EVLNQQTTPFNLUSSBAGWRAT---LLRLG	1225	
Db	69	VWQWVDDALTFELPEIDL-RTNIDPHGTAQALMQADLQO-DLURVDSGKPLVPHQLIOVA	126	
QY	1226	EDDHILTIWVHHITISDGSIVDLRRDLNOLYSAALKSDKPLSLPLP-----IQYSD	1279	
Db	127	DNRWYQRYVHHULVDGFSFPATROIANYICTWLRGEPFPASFPFPADVVEEYQOYRE	186	
QY	1280	FAKWQKQD--FTDEQKOLNWKKQLKSSPAKITPTDFARPALLUSDGACGVHVT-----I	1331	
Db	187	SEAWQRDAFAFWQRQL-----PPASLT-----SPAPLPGRSASADILRLKLEFT	232	

```

1332 DGEYQSLRAFNEHNTTSFVLLAAFAAHYRLTAVEDAVIGTPIANRNRPELEIDIGC 1391
1333 DGEYQSLRAFNEHNTTSFVLLAAFAAHYRLTAVEDAVIGTPIANRNRPELEIDIGC 1391
1334 DGEYQSLRAFNEHNTTSFVLLAAFAAHYRLTAVEDAVIGTPIANRNRPELEIDIGC 1391
233 DGE-FROLATQSGVORTDLAALAAWLG--RLCNRMDYAAAGFIMRRLGSAALATATGP 289
1392 FVNTOCMRINIDHDTFTGLINQVATTTAAENEDIPFVVSALQPGSRDLSSTPLAQ 1451
290 VLNVLPGLGHIAAQETPLATRLAAQLKMRHQRHYDAEQIV--RDSGRAAGDEPLFG 346
1452 LIFAVHSQKDLGRFQGLSEVPVPSKAYTRE---DMEFHLFOETDLSKGSVNAFA--- 1503
347 PVL-----NIKVDYQ-LDIPDVOAHTLATGPNVNDLELALFPD---VHGDLSIEILA 396
1504 -----DELKMETVENVVRVFEILNRGLQSSRTVSLPLPDGIVTLE--KLDVLNVKH 1556
397 NKORYDEPTLIHAERLKMIL-----AQFAADPALLCGDVIDMLPGEYAQLAQLNATQ 449
1557 VDYPRESSLADVFQVQVSAYPSLAVVDSCLRTTETDRQSDILAGWLRRRSMFAETLV 1616
450 VEIP-ETTLISALVAEQAAKTPDAPALADARYLFSYREMRQVVALANLLRGRGVKPGDSV 508
1617 AVFAPRSCETIVAFEGVLKANLAYLPLDVRSRVSARVODILSGLSGPTIVLIGHDTAPPDI 1676
509 AVALPRSVPLTALHAIVEAGAAWLPDITGYDDRLKMLED-ARPSLITITDDQLPREFS 567
1677 EYTNVEFVRIRDALNSNADGFEVIEHDSKPSATSLAYVLYTSGTGRPKGMVIEHRVI 1736
568 DVPNLTSLCYNAPLTPQGSAPLQL-----SQPHHT--AVIIFTSGTGRPKGMVQGTAL 620
1737 IRTVTSGCCIPNYP--SETRMAHMTAIFDGCSEYIYALLFGRTGLVCDVYMTTLDAK 1794
621 VNRLI-WMNHYPPLTGEDVVAQKTPCSEFVSVMFEFFWPFIAKALYMARPEAHRDPLAMQ 679
1795 DYFERHVNAAASHVTSQDVLPRVPRRLSRLTMEFFLVTDSTA-----PD 1841
680 Q-FRAYGVTTTHFVPS-----MLAFAVSLTPTQARQSCATLKQVFCSE 724
1842 ALDA-----QGLYGVOCYNGYGTENGVMSTIYPIDSTE-SFING--VPIGRALNNSG 1892
725 ALPADLCREWQL-TGAPLHNLGYPTAAVDSWYPAFGEELAQVGRSSVPYGPWNWG 783
1893 AVYDPEQQLVIGVWELVVDGLARGYSKA-LDENRFVHTIYNDQTV---RAYRTG 1948
784 LRILDAMHPVPPVAGDLYLFGIOLAQGYLGRPDLTASRF----IADFPAPGERMYRTG 839
1949 DRVRYRIGDIEFFGRMDTQKIRGNRIESAIEAALLR-----DSSVRDAAVLQ--QN 2002
840 DVARW-LDNGAVEYLGSRDDQLKIRCORIEIGEIDRVMOALPDVEQAVTHACVINQAAT 898
2003 EDOAPEILGFVVA 2015
899 GGDARQLVGLVS 911

```

RESULT 11

```

US-09-815-242-13724
; Sequence 13724, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Cart, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078

```

```

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13724
; LENGTH: 1294
; TYPE: PRT
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1294)
; OTHER INFORMATION: Xaa = Any Amino Acid
; US-09-815-242-13724

Query Match      3.3%; Score 538.5; DB 10; Length 1294;
Best Local Similarity 24.5%; Pred. No. 5.6e-30;
Matches 244; Conservative 142; Mismatches 428; Indels 183; Gaps 38;

QY 1114 SGRWFLDQL-DVGSWLYLPIYAVRMGPVNVDAARRMAALAEQRHETLRTTFEDQDG- 1171
DB 10 AQPQIMAEKSLDPLSAWSVAHY-VELNGELDAPLLAKAVAGVQQAQADTLRMFTENG 68
QY 1172 VGQIVHEKLSKEMKVIDLCSGLDLPFEVLNBOPTTF-----NLSSAGNR 1218
DB 69 VQWIDPEHTFGEPPITADL-----RQDPDHLAALMAQADLRQNLRAOSGKP 116
QY 1219 AT---LLRLGEDDHILTIYMHHSIDGWSIDVLRDLNOLYSALKDSDPLSALPLP- 1274
DB 117 LAFHQLIRIDDTRWYQYRHHLLVDGFSFPATITQIAAIIYRAWQSDAPTESPFTPEVD 176
QY 1275 -----IQSDFAKWKQDQFIEQEKQUNYKWKQKQDSSP-----AKIP-----TDARPA 1318
DB 177 VVEYORYROSEAWRDG-----AFWAQORRELPPASMSAAPLPGRSASADILR-M 227
QY 1319 LLSGDAGCVHVTIDGELYOSLRAFCEHNTTSPVLLAAFAAHYRLTAVEDAVIGTPIA 1378
DB 228 KLSAPAGA-----FRLAAMHPEIPRADLALALVTLWLG--RLCGMDYAAAGFIEM 276
QY 1379 NNRPELEIDIGCFVNTQCMRINIDHDTFTGLINQVATTTAAENEDIPFVVS--SA 1436
DB 277 RRMGSAALTATGPNVLPVNLHATDELPTLAKRLAAQLKMRHQRHYDAEQIVRDSG 336
QY 1437 LQPGSRDLSSTPLAQLIFAVHSQKDLGRFQGLSEVPVPSKAYTREDFMEFHLFOETDS- 1495
DB 337 RAAGETPLFGPVLNIKVFYDHY-----LDFFGIAQOHTHTLATGPNVNDLELALFPDENG 389
QY 1496 -----LKGSVNFADELKMETVENVVRVFEILNRGLQSSRTVSLPLPDGIVTLE--- 1546
DB 390 LDIELLANAQRYDDATLSRHALLMALI-----TFADNPA--LRCGDAQMLAAEQ 439
QY 1547 EKLDVLNVKHVDYPPRESSLADVFQVQVSAYPSLAVVDSCLRTTETDRQSDILAGWL 1606
DB 440 TQTLHLNNTAVTIP-AATLSDELVAQQAQKTPESALADAHYHFTYREMRQVVALAYALR 498
QY 1607 RRSMPAETLVAVFAPRSCETIVAFEGVLKANLAYLPLDVRSRVSARVQDILSGLSGPTIVL 1666
DB 499 ERGVQPGDSVAVAPALPRSVFLTLALHGVIEAGAAWLPDITGYDDRLRMML------ 548
QY 1667 IGHDTAPPDIEVTNVEFVRIRDALNSNADGFEVIEHDSKTP--SATSL-----AVY 1716
DB 549 ---EDAQPKLITTTXQALAREHD-----IPGMEYLCYSQPLPVSADATPLGLSLPHHTAYI 600

```


; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059
 ; SOFTWARE: PatentIn ver. 3.0
 ; SEQ ID NO 6480
 ; LENGTH: 1295
 ; TYPE: PR
 ; ORGANISM: Corynebacterium glutamicum
 US-09-738-626-6480

Query Match 2.68; Score 424; DB 9; Length 1295;
 Best Local Similarity 25.8%; Pred. No. 1.3e-21;
 Matches 159; Conservative 112; Mismatches 243; Indels 102; Gaps 25;

Qy	495	EMLKAVSHS-PTKTAIQAGD-WTYSSELDNVSSRLAVHIKSLGLRAQQAIIPIVPEKSK	552
Db	24	DVLESVASTYPEAAID--DQVLTVAELMEEVYALADSIHAQGIKRGDRIGIRMPGSTR	81
Qy	553	WVIASMLAVLKSGNAFTLIDPNPPARTAOVVTQTRATVALTSKLHRETQKLVGRCVV	612
Db	82	DLYTAILATLAAGAAVVPVADDPERAEMVFEANALFDA-----T	125
Qy	613	DDELLQSVASDDFSSLTKSQDLAYVIFTSGTGDGKIMIEHR-----AFSSCALKF	665
Db	126	GFHMLRPTAGD--TRPRDLDTAWIIFTSGTGKPGVAVSHRSAAAFVDAEAQMFVD	183
Qy	666	GASLGINSDFRALQFGTHAFGACELLEIMTLINGGVCVCPISDDDRMNS---IPSFINRYN	722
Db	184	HPSGPIGPEDRVLAGSLVAFDASCEEMWLAHGACL-VPAPRSLVRSGLDGLFWLIRRD	242
Qy	723	VNMMAFPSPYMGTFSPEDVPLATLVLVGEQMSSS-VNAIWAPKQLQLLNGYQSESSI-	780
Db	243	ISVVSTVPTTLAGLWPAEALQVRLIIVGGEACSOELVERLSTPDREVWNTYGPTEATVA	302
Qy	781	CFASNNSTENNMCRAGCAHSWIDPNDRNLVPIGAVGELVIESPCIARDIYVPPPEK	840
Db	303	CGTQLYAGQVGIPLAGWDLVY-VDDAGEPVGIGEVGELVIGVGGLAR-YL---DPEK	357
Qy	841	SPEFTDIPSWYPANTFPDGAKLYRTGLARYADSGSIVCLGRIDSQVKIRGQVRELGAIE	900
Db	358	-----DREKYAPLKS-VGWTRAYRSGDHVLEEDG-LYFVGRVDDQVKIGRRIELGEVD	410
Qy	901	-----THLRQOMPDDLTIVVEATKRSQSANSTSLIAFLIGSSYFGNRPDAHILDHAT	954
Db	411	ANVAALSNVRSS-----AVVVQTT---GADQKVLNAVY-----SLEDAAGAFDHVA	454
Qy	955	KAINIKLEOVLPRHSIPSFICMLELPRATATGIDRRLRMKGKIDLDKOTQGAIVQOAP	1014
Db	455	TA---RUTETMPAALVPRIHV-MDDLPTVTSGKVKDKSL-----P	490
Qy	1015	APIP-----VFADTAAKLHSIWQSLGIDPATVNVGATFELGNGNSITAIKVMNMARS	1067
Db	491	WPLPGTVVEANDLSATEAWIAQEWVDILGTSVSSKD--ADFFSLGGTSLAAATLVGRVA	548
Qy	1068	VGMDLKVSNIYQHPTL 1083	
Db	549	KVPTAAVRDLYDHPRL 564	

Search completed: May 30, 2003, 13:01:36
 Job time : 105 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2003, 12:33:51 ; Search time 39 Seconds
(without alignments)
3327.678 Million cell updates/sec

Title: US-09-482-788-2

Perfect score: 16128

Sequence: 1 MEYLTAVDGRDLPTPTASF.....RVEHLEEVSKTFEGLNSSL 3129

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2084	12.9	6486	1 TYCC_BACBR	O30409 b tyrocidin
2	1975	12.2	6359	1 BACC_BACLI	O68008 b bacitraci
3	1882.5	11.7	5255	1 BACA_BACLI	O68006 b bacitraci
4	1816	11.3	3587	1 TYCB_BACBR	O30408 b tyrocidin
5	1807.5	11.2	2555	1 PPS3_BACSU	P39847 bacillus su
6	1793.5	11.1	3587	1 SRF2_BACSU	Q04747 bacillus su
7	1769.5	11.0	2560	1 PPS2_BACSU	P39846 bacillus su
8	1758	10.9	4451	1 GRSB_BACBR	P14688 b gramicidi
9	1710	10.6	2607	1 BRAB_BACLI	O68007 bacillus li
10	1625.5	10.1	5217	1 HTSL_COCCA	Q01886 cochliobolu
11	1613	10.0	3587	1 SRF1_BACSU	P27206 bacillus su
12	1604	9.9	2561	1 PPS1_BACSU	P39845 bacillus su
13	1515	9.4	3712	1 ACVS_CEPAC	P25464 cephalospor
14	1484.5	9.2	3649	1 ACVS_NOCLA	P27743 nocardia la
15	1422.5	8.8	3770	1 ACVS_EMENI	P27742 emericella
16	1347.5	8.4	3791	1 ACVT_PENCH	P26046 penicillium
17	1346.5	8.3	3746	1 ACVS_PENCH	P19787 penicillium
18	958	5.9	3947	1 STD2_USTMA	O43103 ustilago ma
19	909.5	5.6	1278	1 DBFB_BACSU	P45745 bacillus su
20	852	5.3	4447	1 PRKS_BACSU	P40803 bacillus su
21	756	4.7	1274	1 SRF3_BACSU	Q08787 bacillus su
22	742	4.6	1098	1 GRSB_BACBR	P14687 bacillus br
23	740	4.6	1088	1 TYCA_BACBR	P09095 bacillus br
24	654.5	4.1	2035	1 HMP2_YEREN	P48633 yersinia en
25	568	3.5	1293	1 ENTF_ECO57	O8xbv9 escherichia
26	559	3.5	1293	1 ENTF_ECOLI	P11454 escherichia
27	514	3.2	1419	1 LYS2_SCHFO	P40976 schizosacch
28	494.5	3.1	1392	1 LYS2_YEAST	P07702 saccharomyc
29	469	2.9	805	1 ACVS_STRLC	Q01757 streptomyce
30	451	2.8	1048	1 ANGR_VIBAN	P19828 vibrio anqu
31	434.5	2.7	1391	1 LYS2_CANAL	Q12572 candida alb
32	408.5	2.5	503	1 DUTA_BACSU	P39581 bacillus su
33	383	2.4	506	1 DUTA_LACCA	P35854 lactobacilli

ALIGNMENTS

RESULT 1

ID	TYCC_BACBR	STANDARD;	PRT;	6486 AA.
AC	O30409;			
DT	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Tyrocidine synthetase III [Includes: ATP-dependent asparagine			
DE	adenylase (AsnA) (Asparagine activase); ATP-dependent glutamine			
DE	adenylase (GlnA) (Glutamine activase); ATP-dependent tyrosine			
DE	adenylase (Tyra) (Tyrosine activase); ATP-dependent valine adenylase			
DE	(ValA) (Valine activase); ATP-dependent ornithine adenylase (OrnA)			
DE	(Ornithine activase); ATP-dependent leucine adenylase (LeuA) (Leucine			
DE	activase)]			
GN	TYCC.			
OS	Bacillus brevis.			
OC	Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.			
OX	NCBI_TaxID=1393;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 8185;			
RX	MEDLINE=98012987; PubMed=9352938;			
RA	Mootz H.D., Marahel M.A.;			
RT	"The tyrocidine biosynthesis operon of Bacillus brevis: Complete			
RT	nucleotide sequence and biochemical characterization of functional			
RT	internal adenylation domains.";			
RL	J. Bacteriol. 179:6843-6850(1997).			
CC	-!- FUNCTION: INCORPORATES SIX AMINO ACIDS (FOR TYROCIDINE A, ASN,			
CC	GLN, TYR, VAL, ORN, AND LEU) IN THEIR L-CONFIGURATION INTO THE			
CC	PEPTIDE PRODUCT.			
CC	-!- COFACTOR: CONTAINS 6 COVALENTLY BOUND PHOSPHOTETRAINES (BY			
CC	SIMILARITY).			
CC	-!- PATHWAY: Cyclic peptide antibiotic tyrocidine biosynthesis.			
CC	-!- SUBUNIT: LARGE MULTIMERIC COMPLEX OF TYCA, TYCB AND TYCC.			
CC	-!- DOMAIN: CONSISTS OF SIX MODULES, AND HARBOURS A PUTATIVE			
CC	THIOESTERASE DOMAIN AT ITS C-TERMINAL END. EACH MODULE			
CC	INCORPORATES ONE AMINO ACID INTO THE PEPTIDE PRODUCT AND CAN BE			
CC	FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR SUBSTRATE-			
CC	ADENYLATION, THIOLEATION, CONDENSATION (NOT FOR THE INITIATION			
CC	MODULE), AND EPIMERIZATION (OPTIONAL), AND N METHYLATION			
CC	(OPTIONAL).			
CC	-!- MISCELLANEOUS: TYROCIDINE IS A MIXTURE OF FOUR CYCLIC			
CC	DECAPEPTIDES; TYROCIDINE A (D-PHE-PRO-PHE-D-PHE-ASN-GLN-TYR-VAL-			
CC	ORN-LEU), B, C, AND D, IN WHICH PHE, AT POSITIONS 3, 4, AND TYR			
CC	RESIDUES ARE GRADUALLY REPLACED BY TRP, DEPENDING ON THE RELATIVE			
CC	CONCENTRATIONS OF THESE AMINO ACIDS IN THE GROWTH MEDIUM.			
CC	-!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME			
CC	FAMILY.			
CC	-!- SIMILARITY: CONTAINS 6 ACYL CARRIER DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			

or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL; AF040835; AAC45930.1; -
DR HSSP; P14687; 1AMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatin.
DR InterPro; IPR003880; Phantne.attach.
DR InterPro; IPR000379; Ser_estrs.site.
DR InterPro; IPR001031; Thioesterase.
DR pfam; PF00501; AMP-binding; 6.
DR pfam; PF00550; pp-binding; 6.
DR pfam; PF00668; Condensation; 6.
DR pfam; PF00975; Thioesterase; 1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 6.
DR PROSITE; PS00455; AMP_BINDING; 6.
DR PROSITE; PS00075; ACP_DOMAIN; 6.
KW Ligase; Antibiotic biosynthesis; Phosphopantetheine;
KW Multifunctional enzyme; Repeat.
FT REPEAT 466 1038
FT REPEAT 1521 2070
FT REPEAT 2536 3113
FT REPEAT 3590 4149
FT REPEAT 4606 5203
FT REPEAT 5658 6245
FT REPEAT 970 1037
FT DOMAIN 2007 2074
FT DOMAIN 3045 3112
FT DOMAIN 4080 4147
FT DOMAIN 5124 5191
FT DOMAIN 6167 6234
FT BINDING 1000 1000
FT BINDING 2037 2037
FT BINDING 3075 3075
FT BINDING 4110 4110
FT BINDING 5154 5154
FT BINDING 6197 6197
FT BINDING 6486 6486
FT SEQUENCE 4334900AF07DF786 CRC64;

Query Match
Best Local Similarity 23.0%; Score 2084; DB 1; Length 6486;
Matches 864; Conservative 595; Mismatches 1372; Indels 928; Gaps 136;

QY 46 IEAKPCTPQDMDICNALDKQSAIGHAYD-----VPTDIDSRFALANK 92
DB 7 IAKIYPLTLQ-----EGMLFHAVTDTGSSAYCLQMSATIEGDFHLPFKSLN 55
QY 93 EIVNQTALRAFAFTSDGKTSVILKDSFVFWMCWSSSSSDPEVVVDEAAAAAGPRC 152
DB 56 KLVENYVLTAFVYQNMQRPRVFKERKV-TVPCENTAHLP--AEQDAYIQAYTKQH 112
QY 153 NRVLLD-----MGT--KKQQLVWTFSHALVDVTFQOVLRSVFAAYKKHEKDTHRPE 203
DB 113 HAPDLTKDNLKMAAIFQTAENKYRLVWAFHHIIVDGTGLVLLKLLTYVA-----ALRKG 168
QY 204 TPSSDATDTSQSVSWSCEDNAVSAFTHQWTHNDLN-----ASVFPH 250
DB 169 EPIPREATPYSEYIKWLDKQNKDEALA---YQNYLAGYDHOAAFPKKKLGTEASRYEH 225
QY 251 LSDHLVMPNTTAEHRITFPLSOKALSNAICRTALSTILLSRYTHSDALFGAVTEQSL 310
DB 226 VEAMPTIAPEKTOQLQIA---NONQATSSVFOALWGLASTYKKADDDVFGSVVSGRP 282
QY 311 P-FDKHYLDAGTYQTVAPLRVHCQSNLRASDMDAI-----SSYDRLGLHLPGLRD 362
DB 283 PQIQIESWGLVPIIPTPRVQTNKQOTSELLQTVQKQALASATVD-----FAP--LYE 335
QY 363 IRNTGDSNACDFQVL-----LVTDGSHVNNNGINGFLQOITESSHF-----406
DB 336 IQST-----TVLQKELIDHLVTFENYPDHS-----MKHLEESLGFQFTVESGDEQ 380
QY 407 -----PCNNRALLHCOMESSGALLVAYYDHNVIDSLQTTLLQOFGHLIKCL 454

```

```

DB 381 TSVDLNVVALAPSNE-----LYVKLSYNAV-----YESSFVNRIE-----GHLRTVI 424
QY 455 QSPDLDS--SMAEVNLMTEYDRAE-IESWNSQPLEV--QDTLIHHEMLKAVSHSPTTAIQ 510
DB 425 DQVIGNPHVILHEIGIITEEEKQOLLVAYNDTAEYPRDKTIFELTAEQASRTPAKAAV 484
QY 511 AMDGWTYSELDQNVSRRLAVHISGLGRQAQAIIPVYFEKSKWVIASMLAVLKSNGAFTL 570
DB 485 CGEDTLTYQELMERSAQLANALREKGI-ASGSIVSIAEHSLELIIVAIMAVLRSGAAYLP 543
QY 571 IDPNDPPARTAQVVVTRATVALTSLKHLRETQVOKLVGRVGVVDELLQSVASDDSSLT 630
DB 544 IDPEYQDRIQYLLDDSQTTLLLTQS-HLQPNTRFAGSVLYLDDRSLYE--GGSTSFAPES 601
QY 631 KSODLAYVFTSGTGDPRGIMIEHRAFSCALKFGASLGINDTRALQFGTH---AFGA 687
DB 602 KPDDLAYMIYTSSTGTGNPKGAMITHQGLVN-YIWANKVYVQGE--AVDFPLYSISFDL 658
QY 688 CLLEIMTTLINGCGVCIPSDDDRMNSIPSEFINRYNNMMATPSYMGTFSPEDVPLATL 747
DB 659 TVTSIFTPLLSGNTIHVYRGADKVQVILDIKDKNGVGIKLTPTHLKLI--EHIDGKASS 716
QY 748 V---LVG-----BOMSSSVNAIWAPKLQLLNGYGQSESSICFA---SNMSTEPNNNGRA 796
DB 717 IRRFIVGGENLPTKLAKIYDHEGENVQIFNEVGPTETVVGCMYIYLDPTQTQESVPIG 776
QY 797 VGAHS---WVIDPNDINRLVPIGAVGELVIESPGIARDYIVPPPEKSPFTDIPSWYPA 853
DB 777 VPADNVOLVLLDAS--MQPVPVGLSEMYIAGDQVAKGYENRPELTKEKID-----826
QY 854 NTFPDCAKLYRTGDLARYASDGSIVCLGRIDSQVKGORVELGAIETHLRQ--OMPD DLT 912
DB 827 NPERPGTKMYRTGDLAKWLPDGNMEYAGRMVQVKGIRHRIEMGEIETRLTQHEAVKEAV 886
QY 913 IVVEATKRSQSANSTSLIAFLIGSSYFGNRPNSDAHILDDHDKA--INIKLEQVLPKHSIP 971
DB 887 VIVEKDESGQNV---LYAYLVS-----ERELTVAERREFLGRITLPSYMIIP 928
QY 972 SFYICMLELPRTATGKIDRRRLRMKGDLDKOTGAIVQOAPAPIPVFADTAALKHSIW 1031
DB 929 SFPIRLAEIPLTANGKVERKKL-----PKPAGAVVTGTAYAAPO-NEIEAKLAEIW 978
QY 1032 VQSLGIDPATVNVGATFFELGNGSITAIKMW-NMARSVGNLDKLVSNYIQTHTLAGISAVV 1090
DB 979 QQVVLGI--SOVGTHDDFDLGGHSLKAMTVVFOVSKALEVELPVKALFEHTVVAELARFL 1036
QY 1091 -KGDPLSYTLI-PKSTHE-GPVEQSYSGRLWFLDQLDVGSLWYLPYAVMRGPNVDA 1147
DB 1037 SRSEKTEYTAIOPVAQEEYVP--SSAQKEMYILQOFEENGISYNTISGAILLEGKLDYAR 1094
QY 1148 LRRALAEQRHETLRTTFEDQDQGVQIVHEKLSBEMKVIDLCGSDLPDFEVLNQEQT 1207
DB 1095 FASAVQQLAERHEALRTSFHRIDGEPVQKVHEEVEVPLFMLE---APEQAEKIMREFVR 1151
QY 1208 PNLSEAGWRATLLRLGEDDHLITIVMHHIISDGSIDVLRRLDQLNQLISAAALKDSKDL 1267
DB 1152 PFDLGVAPLMRTGLLKGDKRHLFLDDMHIIISDGVSSQILLREFAELOYGA-----1203
QY 1268 SALTPLEIOYSDPAKWQKQOFIEQ--EKOLNYKKQLKSSP-AKIPDFARPALLSGDA 1324
DB 1204 -DLQPLSLQYKQFAANQNELFQTEAYKKQEQHWLNTFADEIPLNLPTDYPFVSQSFAG 1262
QY 1325 GCVVHTIDGELXQSLRAFCEHNTTSFVLLAAAFRAAHYRLTAVEDAVICTPIANRRPE 1384
DB 1263 DLVLFAGKELLERLQOVASETGTLTYMILLAAYNVLLSKYTQGEDIVGTPVAGRSHAD 1322
QY 1385 LEDIICCFVNTQCMRINIDHDDHDFGLINQVKTATTAAFENEDIPERVVVSALQPGSRDL 1444
DB 1323 VENIMGIFVNTALRNOPASSKTFQAFLOEVKQNALAAVDHODYPPEELVEKLAIAI-ORDI 1381
QY 1445 SSTPQAQLFAVHSQDKGRFKFGLESVPVPSKAY-----TRFDMEFHLFQETDLSKG 1498
DB 1382 SRNPLFDTLSLENANQ-----QSLAIAELTASPYELFNKISKFDLALNASESPADIQF 1435

```


Db 496 QIARLLREKGI--GRSIAAIIADRSEFEMIIGIIGILKAGGAYLPIDPETPKHRIAPMLSD 554
QY 587 TRATVALTSKLHRETQVQLVGRVVDDELQSVASDDF-----SSLTKSODLAVYFT 641
Db 555 TKAGVLLAQGAADGID-----CEA--DIIHLDKGVADGFSKKRLSSVNDSDGTAYIYT 607
QY 642 SGSTGDPKGIIMEHRAFSCALKFGASGINSIDTRALOGTHAFGACLEIMTTLINGCC 701
Db 608 SGSTGMPKGVVTPHYS--AARVYKNTNYIDITEDDAILQLSNYSFSGSVDFIDFALLNGAS 666
QY 702 WCIPSDDDORMNS--IPSEFNRYNVNMWMTAPSYMCTFSPEDVPGLATL---VLVGEOMS- 755
Db 667 LVLIKEKTVNLHAEVETKREQSVSMFTTALFNTLADINIGCLAKURKIFLGGERASI 726
QY 756 ----SSVNAIWAIPKQLLNGYQSES-----SSTCFASNMSTENNMGRAGV---AHSWVID 805
Db 727 PHVRKVLNVRGRDK--LIHVGYPTSTVYATVYFINEIDDEAETI--PIGSPPLANTSULI 782
QY 806 PNDINRLVPIGAVGELVETSPGIARDYI-----VPPPEKSPFTTDPSPWYPANT 855
Db 783 MDEAGKLLPIGVPGELCIAGDGLSGYLNREBELTAEKFIHP-----824
QY 856 PDGAKLYRTGLARYASDVCLGRIDSQVVKIRGORVELGAIETHLRQMPDDL--TI 913
Db 825 FIPGERLYKTGLAKWLPDGNIEFGRIDHQVIRKGFRIELGEIESRL--EMHEDINETI 882
QY 914 VVEATKRSOSANSTSLIAPLIGSSYFGNRPSDAHILHDHATKAINIKLEQVLPRIHSPF 973
Db 883 V---TVREDEESRPYICAVIT-----ANRE-----ISLDELKGF---LGEKLPYEMIPAY 926
QY 974 YICMLELPTATGKIDRRRLRMGKDILDKOTQOGALVQ--QAPAPIVFADTAAKLHSIW 1032
Db 927 FYKMDKLLPTKNGKVDKALP-----EPDRTAGAENEYEAPR-----NETEKLAAIWR 975
QY 1033 QSLGIDPATVNVGATFFELGGNSITAIRM--VNARSVGMDLKAVSNLYOHPHTLAGISAVVK 1091
Db 976 DILKVEKSGIN--DIEFFEMGHSLKAAAMAARIKELKAEIPLGQIFKPTTKGLEYIR 1033
QY 1092 GDPLS--YTLIPKSTHEGPVEQYSQGRFLWFLDOLDVGLWYLIPIYAVRMGRPVNVDALRR 1150
Db 1034 STKDSVYSSIOKVEEKEYRLSSAQRKLYILDQIEGSLGSLNIPFTMKVKGRDIRRFEN 1093
QY 1151 ALAALQORHETLRTTFEDODGVGVQIVHEKLEEMKVIDLCSDDLDPFVLNQEQTTTPFN 1210
Db 1094 ALKTTIORHEALRTFLMADGEPVQKIEKVEFDSIKCKSI--QSLSIQIEIIRKQ--FVRPFD 1150
QY 1211 LSSEAGWRATLRLGEDDHILTVMHIIISDCWSIDVLRDLNOLYSAALKDKDPLSAL 1270
Db 1151 LKAPLFRTEVVKVDDDEHIIILFDMHIIISDASGMLTKEICDLYG-----KEL 1201
QY 1271 TPLPIQYSDFAKQKD--QFIEQKQNLNWKQKLDSSPA--KIPTDFAFPALLSGDAGCV 1327
Db 1202 KPLSLQYKDYSEWQDFYQKDEMKQKEYLWLFKEIPLVLMNPTDYPRQMHVSBEGRDI 1261
QY 1328 HVTIDGELYQSIRACNEHNTTFSVVLLAFAAFAHRLTAVEDAVIGTGTPIANRNPELED 1387
Db 1262 GRAIDGELTKKLRIAKDNGATMYMLLAAYTVLLRTYSQGEVDVIITGIQGRKHHELKH 1321
QY 1388 IIGCFVNTOCMRINDHDDFTGTLINOVKATTAAAFENEDIPFERVVSALQPCSRDLST 1447
Db 1322 VLGMEVNTLAMRNHPKGTFFAYEQDVKETALKAYENQDYQFDOLVEQLDL--ERDMSRN 1380
QY 1448 PLAQLIFAVHSOKDLGRFQFQLESVPVPSKAY--TRFDMEFHLFQETSDLSKGSVNFADL 1506
Db 1381 PLFDMTFLVQLNEKADA--BIEGLTEPPFESDIHISKFDLTLAIEKDKSIEFDLEYCTKL 1439
QY 1507 FMNETENVVRVFFELRNLGLOSRTPVSIPLDTGIVTL-----EKLDVL---NVKHVDYVP 1560
Db 1440 FRETVERMAAHFVRVLED--ISKRTDKRL---DQIEAMSEDEKNTLLYRFENDTKTDAP 1493
QY 1561 RESSLADVPQTVQSAYPDSLAVVSDSCRLTYTTELDRQSDILAGWLRRRSPAEITLVAVFA 1620
Db 1494 TDKTIQULFAERAETSPDKTAVVFEDQTLTYRQLHRSNQLARFUREKGVQPDATVGINV 1553

QY 1621 PRSCETIVAFFGVGLKANLAYLPDVRSPSARVQDILSGLSGPTIVLIGHDHTAPDIEVTN 1680
Db 1554 DRSPEMIGLGLGILKAGGAYLPDPAYPEDRIKYILG--DSQTKFLLSEAL-----IKK 1606
QY 1681 VEFVRIRDALNSDNGDFEVIHDSHTKPSATS-----LAVLYTSGTGPKPKGVMIEHRVI 1736
Db 1607 RAFIREKADINDIHDK--QIAAQDAQAQLEFPVSRSGDLAYIITSGTSGRPGKGLVIRKGL 1665
QY 1737 IRTVTSGC--IPNVPSETRMAHMATIAFDGASVEIYSALLFGRTLVCDVYMTTLLDARALD 1795
Db 1666 SNLVSAAVKMLHNLGSRVIOFASLSFDASAEIFFPALAAGSALVLRGEBEMPGOPLRS 1725
QY 1796 VEFREHVAASHVTSSQDVPLRPRRLSRTLMFFVFLVYVTDSTAPDALDAQGLVQGVQCY 1855
Db 1726 -FLROY--NITHAT-----LPPTVLDVNLSEGLNKLKIVVSAGSACSEELAKRWSGNRLF 1777
QY 1856 -NGYPTENGVMST--IYPIDSTESFINGVP--IGRALNNGSAYVDPBQQLVIGVYMGEL 1911
Db 1778 INAYGPTETVTCATAGIYE-----GSRPHIGSPIANTNVYLDONKQKVPFTGVVGL 1830
QY 1912 VVTGDLARGYSK--ALDENRFV--HITVNDQTVKAYRTGDRVRYRGDGLIEFFGRMDTQ 1969
Db 1831 CVGNSLARGYLNRRPELTAEKFIHPFASGE--RLYRTGDLARW--LPDGHLEPLGRIDHQ 1887
QY 1970 FKIRGNRIESAEIEAALLRDSSVYRDAAVVYLQONEDQAPFELGVFVADHDHSENDKQGSAN 2029
Db 1888 VKIRGYRIELGEIENOLLKLDKIDEAAVIAARKDDHSDVLCAYIVSKED-----1936
QY 2030 QVEGWQDHFESGMYSDIGELDPSTIGSDFKGTSMYDQSGIDFEDENHWEHGEITRTHLDN 2089
Db 1937 -----WTS-----TEISEW-----1945
QY 2090 RSLGNVLEITGTGSMILFNLDLSRLESYVGLFSPSAAAFVFNKATESIPSLAGKAKVQVGT 2149
Db 1946 -----LEKELPHYM-----IPAYFVR--LDKLP-----1966
QY 2150 ATDIGOVDL-----HPDLVVLNSVIOYFPSPSEYLAETADTLIHLPNVOR--IPFGDVRVSOA 2204
Db 1967 LTSNDKVRKALPAPDRHVATGAVYEAPRNDTEAKLVD-----IWRDVLGAGDI---G 2016
QY 2205 TNEHFLAA-----RAHTLGKNATKDDVRQKMA--ELEDMEBELLVPEAPFTSLKDR 2254
Db 2017 ISHFFFAAGDSITAKQIVSRUSRLGKLEMKDOLFANPRIKDLAKYVKKO-----SORKNA 2072
QY 2255 FPLGVHEVETLP-----KNMEAVNELSAYRYAAVHVVRGSLGDELVLPL-----VEKDD 2302
Db 2073 NTIVTGHAEITPQKWYFANNKE--ELDHFNQSFVLFRKGGFDESCVKKAFNKIMEQHD 2129
QY 2303 WI-----DF-QANQLNQKSLGDLKSSDAAIMAVSKIPPEITAFERQV--VASLNS 2350
Db 2130 ALRMIYEEKGGDFIQYNRGFRDLDL---DV-----YDVRGLDRQAEKVVELAT 2176
QY 2351 NIDEWOLSTIRS-----SAGDSSLSVPDIFRIAGEAFRVEVSSAROWSONGAL 2400
Db 2177 SIQ--KLSSIRKGLVHLGIFRADEGHLLIVLHLLVVDG--VSWRI-----L 2220
QY 2401 DAVFHHCCSO--GRTL--VNFPTDHLRGLDILLTNRLQRLQNRRIAIEVRRLRSLPLS 2456
Db 2221 FEDFETLYSQALKGOTLEICYKTDSYQEEFA-----RLKAYAHSRITLSKAEYWRNIKA 2275
QY 2457 ---YMTIPSINIVLDKMLNANGKVRKELSRRAKVVPKQOOTAAPLTPFISEVEVILCEE 2513
Db 2276 RVRFIPKKNVL-----KEDVYENS-----TTLSIKLGE 2304
QY 2514 ATEVFGMKVDI---TDHFFNLGCHSLLATKLISRI-----DQRLKVRIT---VKDVPD-- 2560
Db 2305 AT-----ADLLRNTNRAYNTEINDILLTALTGARDITGENKLVKMMEGHGEDILEGV 2358
QY 2561 -----HPVFADLA--SVIROGL-----GLQOPVSDGOGDRSAHAPRTETEA 2601
Db 2359 DITRTIGWFTMYPVLLDAGEEAKLSQIKMWKETLRKIPNKGIGYGLLKLYMA---EDPD 2415

QY 2602 ILDEFKAVLGQVIGITDNFEDLGGHSLMATKLAVRIGHRLDTTVSVK----DVFDPHPVLV 2658
Db 2416 FTNEEKARISFNVLGDIADQWNGEFGSGSFSEGISGGRIARSHSIEINAIYVMHSLVI 2475
QY 2659 -----QLATALDNLVSKTNEIVGGREMAEYSPFQLLFTEPPEFM 2699
Db 2476 HTTFNQMEYKEDTISRLNHLQKLEQIIKHCQTQTESERTPSYGDGTNLSLAE----- 2529
QY 2700 ASEIKPOLEOIIQDIYPTQMKAPLF-----DHTT-----ARPRPVFPYIDFPSTSEP 2751
Db 2530 LEEIKG--KYRSAIEKTYPLANQKGMFLFAIEDHTSDAYFQQTVMIEGYVD-PAILEA 2586
QY 2752 DAAGLACASLNVHLIDFTVFAEASGELYQVVLSCLDLPQIVETDNI-----NTAT 2806
Db 2587 SFNDIMKRHEIL-----RASYE-YEIV-----EPRQIIENRSDIFTYFNIAK 2629
QY 2807 NE-----FUDFAKEPVRGLHPLIRTIK-----QTKSMRVIMRISHALYDGLSLH 2854
Db 2630 SSAQQQOEFTLRLNEDRKKGFDLSKDVLMRAYLLKTAERSYRLVSHHHLLDGCWGLGI 2689
QY 2855 VVRKLHMLY-----NGRS--LLPPHQFSRYMOY-TADGRESGHGFWRDVI-----QNTPTMTI 2903
Db 2690 IMRELFVIYENRMNGRASPLKPKYSYIKWLEODQOEAREQYRWREYLKGYEEQAOLPT 2749
QY 2904 LSDDTVDGNDATCKALHLSKIYNIPSOVLGRSSNIITQATVFNACALVLSRESKDV 2963
Db 2750 LTRKKSSRYDRREKVHLSKQTKQLKEL-AARNSVTLHTVTOTIWLMLTRYTKIDVV 2808
QY 2964 VFGRIYSGROGLPVEQDIYVGPCTNAVPRRAHI-ESSDYNOLLHDIDQVLLSLPH---- 3018
Db 2809 VFGTVYSGREANVDGIEDMIGLFINITPIRNEQARFNDCLQKVEDAIOQSNRYNMN 2868
QY 3019 --ETIGFSDLKRNCTDMPETATNFSCCITYHNFEYIPESQFEQORVMGV----- 3066
Db 2869 LAEQVALLSKKLIDH-----ILVFNVE-ADEQDFESSQMKTKGKVEISAEEQ 2918
QY 3067 -LTKF-----VNIEMDEPLVDLATAAGVEPDGAGLKYTVIAKTOLFGRK--RVE 3112
Db 2919 SITAFMSVTPGELFLVLYDGNVDRDIINNIE---GHKIRVAEQVTANENRKAIEID 2975
QY 3113 HLLEEVSKT 3121
Db 2976 MIAEERKT 2984

RESULT 3

BACA_BACLI STANDARD; PRT; 5255 AA.
AC O68006;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE Bacitracin synthetase 1 (BAC1) [includes: ATP-dependent isoleucine
DE adenylation (ileA) (isoleucine activase); ATP-dependent cysteine
DE adenylation (cysA) (cysteine activase); ATP-dependent leucine adenylation
DE (leuA) (leucine activase); ATP-dependent glutamate adenylation (ileA)
DE (glutamate activase); ATP-dependent isoleucine adenylation (ileA)
DE (isoleucine activase); Glutamate racemase (EC 5.1.1.3)].
GN BACA.
OS Bacillus licheniformis.
OX Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OC NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10716;
RX MEDLINE=98089193; PubMed=9427658;
RA Konz D., Kleus A., Schoegeendorfer K., Marahiel M.A.;
RT "The bacitracin biosynthesis operon of Bacillus licheniformis ATCC
RT 10716: molecular characterization of three multi-modular peptide
RT synthetases";
RL Chem. Biol. 4:927-937(1997).
CC -1- FUNCTION: ACTIVATES FIVE AMINO ACIDS, INCORPORATES TWO D-AMINO
CC ACIDS, RELEASES AND CYCLIZES THE MATURE BACITRACIN.

CC -1- CATALYTIC ACTIVITY: L-glutamate -> D-glutamate.
CC -1- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTHETHEINES
CC (POTENTIAL).
CC -1- PATHWAY: Cyclic peptide antibiotic bacitracin biosynthesis.
CC -1- SUBUNIT: LARGE MULTISUBUNIT COMPLEX OF BAI, BA2 AND BA3.
CC -1- DOMAIN: CONSISTS OF FIVE MODULES AND ONE EPIMERIZATION DOMAIN IN
CC THE FOURTH MODULE. EACH MODULE INCORPORATES ONE AMINO ACID INTO
CC THE PEPTIDE PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS
CC RESPONSIBLE FOR SUBSTRATE ADENYLATION, THIOLATION, CONDENSATION
CC (NOT FOR THE INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND
CC N METHYLATION (OPTIONAL).
CC -1- MISCELLANEOUS: BACITRACIN IS A MIXTURE OF AT LEAST TEN CYCLIC
CC DODECAPETIDES, THAT DIFFER BY ONE OR TWO AMINO ACIDS. THE MOST
CC ABUNDANT IS BACITRACIN A, A BRANCHED CYCLIC DODECAPETIDE. IT
CC CONTAINS AN N-TERMINAL LINEAR PENTAPEPTIDE MOIETY (ILE-CYS-LEU-D-
CC GLU-ILE) WITH AN ISOLEUCINE-CYSTEINE THIAZOLINE CONDENSATION
CC PRODUCT AND A C-TERMINAL HEPTAPEPTIDE RING (LYS-D-ORN-ILE-D-PHE-
CC HIS-D-ASP-ASN), IN WHICH THE FREE ALPHA-CARBOXY GROUP OF THE C-
CC TERMINAL ASN IS BOUND TO THE EPSILON-AMINO GROUP OF LYSINE. IT
CC CONTAINS FOUR AMINO ACIDS IN THE D-CONFIGURATION (GLU-4, ORN-7,
CC PHE-9, AND ASP-11).
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF007865; AAC06346.1; .
CC HSP; P14687; IAMU.
CC InterPro: IPR000873; AMP-bind.
CC InterPro: IPR001242; Condensatn.
CC InterPro: IPR003880; Ppantne-attach.
CC Pfam; PF00501; AMP-binding; 5.
CC Pfam; PF00550; pp-binding; 5.
CC Pfam; PF00668; Condensation; 5.
CC PRINTS; PR00154; AMPBINDING.
CC PROSITE; PS00012; PHOSPHOPANTHETHEINE; 3.
CC PROSITE; PS00455; AMP-BINDING; 5.
CC PROSITE; PSS0075; ACP_DOMAIN; 5.
CC Ligase; Isomerase; Antibiotic biosynthesis; Phosphopantetheine;
CC Multifunctional enzyme; Repeat.
CC REPEAT 39 612
CC REPEAT 1109 1648 DOMAIN 1 (ISOLEUCINE-ACTIVATING).
CC REPEAT 2124 2689 DOMAIN 2 (CYSTEINE-ACTIVATING).
CC REPEAT 3164 3732 DOMAIN 3 (LEUCINE-ACTIVATING).
CC REPEAT 4668 5249 DOMAIN 4 (GLUTAMINE-ACTIVATING).
CC DOMAIN 621 1037 CYCLIZATION (POTENTIAL).
CC DOMAIN 544 611 ACYL CARRIER (ACP) 1.
CC DOMAIN 1585 1652 ACYL CARRIER (ACP) 2.
CC DOMAIN 2621 2688 ACYL CARRIER (ACP) 3.
CC DOMAIN 3664 3730 ACYL CARRIER (ACP) 4.
CC DOMAIN 5171 5238 ACYL CARRIER (ACP) 5.
CC BINDING 574 574 PHOSPHOPANTHETHEINE (BY SIMILARITY).
CC BINDING 1615 1615 PHOSPHOPANTHETHEINE (BY SIMILARITY).
CC BINDING 2651 2651 PHOSPHOPANTHETHEINE (BY SIMILARITY).
CC BINDING 3694 3694 PHOSPHOPANTHETHEINE (BY SIMILARITY).
CC BINDING 5201 5201 PHOSPHOPANTHETHEINE (BY SIMILARITY).
CC SEQUENCE 5255 AA; 598254 MW; 906E8DD68450F85B CRC64;

Query Match

Best Local Similarity 11.7%; Score 1882.5; DB 1; Length 5255;
Matches 777; Conservative 603; Mismatches 1334; Indels 637; Gaps 131;
QY 41 LDSRRIEATKPTPTFOLDMDICNDKQSAIGHAVYDVPT-----DIDISRFALAKWEI 94
Db 1663 IQAREKEYTPTPAQORMYMLSENE-----RCAYHIPMALLIVEGRINAMQLENALKTF 1718

Db 3654 GTGYD-----PPRNEIERKLVQVWREILGAEIDIGISHHFFAAGDSKALQIVSRL-- 3704
QY 2198 GDVRSQATNEHFLAARAIHTLGNATKDDVROKMAELEMEDEELL-VEPAFTSLKDRFP 2256
Db 3705 AKNNLKEMKALFANPKIDLSRFEITEETHRKNRPVVTGETETELPIQKRYFANNKEELD 3764
QY 2257 GLVHEVILPKMEAVNEL-SAVRYAAVHVVRGSLGDELVLPEKDDWIDFQANLNQKS 2315
Db 3765 HFNQSFMLFRKGDYDENIVRTAFNKILEQH-----DALRMIYEEDGDGTIOYNRGYREN 3818
QY 2316 LGDLLKSSDAAINAVSKIPETITAFERQVVASLNSNIDWOLSTIRSSAEGDSSLSVPDI 2375
Db 3819 LFLD-----DV-----YDVRGFSQ-----EKKVPELATGQKSSSI----- 3950
QY 2376 FRIAGEAGREVESSAROWSQNGALDAVPHHCCSQGRTLVNFPDTHHLRGSLLTNRLPQ 2435
Db 3851 -----RKGKLVHLGIFRADEGDHLLTAI-HHLVVDG----- 3880
QY 2436 RLQNRRIATREVRRLSLPSYMPISNIVVLDKMPNANGKVDK-KELSRRAKVVPKQ-- 2493
Db 3881 --VSWRILFEDFETL-----YLAQKGEPLDIGYKTSYQEFARQLKKAQSR 3927
QY 2494 -----TAAPLTPPISBEVILCE-EATEVFGMKVDITDHF-----FNLGGH 2534
Db 3928 LKEREYWKALEADVPFPAEKLERDTEHSATLSIRIGPDVTAKLLRNAFKAYNTEIN 3987
QY 2535 SLATKLISRIDQLKRVITKVDVDFHPFADLASVIRQGLGQVPDQSGQDQSAHMA 2594
Db 3988 DILLTALIA-----AVRDIITGE-----NKLKVMMECHG-REDILDGV----- 4023
QY 2595 PRTEATAILCDEPAKVLGVQVGTIDNFFDLG-----GHSMLATKLAVER-----IGH--- 2640
Db 4024 -----DITRIGFTVTPYVFDLGEKEISQNKVMYKEALRIPNGIGYVL 4072
QY 2641 RLDTTVSVKDVDFHPVLFQALALDNVQSKT-----NEIVGGR----- 2679
Db 4073 KYMTEELQKIQTAPLSFNFYFGEMNDNRKVFSPFSGESIGGKIYRHCALMNAIS 4132
QY 2680 ---EMAEYSPP-----QL--LFTEDPEFM-----ASEIKQ-----LE 2708
Db 4133 LAGELTIYTFNODQYQTSITIEOLNGSFENLEKIVDHCVDKESDMPDSDYGDVSLGLE 4192
QY 2709 LQEIIOD-----IYPTOMQKALF-----DHTTARPRPVFPVDFPSPSEPDAG 2755
Db 4193 ELELIKDKYSAFQIEKIYPLANNKGMFLFNAMDQISG--AYQOIVIKLGRVHDI-- 4248
QY 2756 LKACESLNHLDFRTVFA-EASGELYQVLSCLDPLQIOVIEDNINATNEFLDEFA 2814
Db 4249 LEESFHEIVKRHEILRASFEYEITAEPROIARDKTPFTSIDLTGENRTRQHRFIETVL 4308
QY 2815 KEPVRLG-----HPLIRFTIIKOT-KSMRVIMRISHALVDGLSLHVRVKLHMLY----- 2863
Db 4309 KEOERKGFDSLSSPALMRVCLIKMSDESRYRLIWSHHILLDGCWGLGIVLSFLYKIMK 4368
QY 2864 -NGRSLPLPHOFESRYMOY-TADGRESGHGFWRDVION-----TPMTILSD 2907
Db 4369 GESRRLKEPKPYGDYIKWLEKQDOEBAYKWKDYLKGYRSRSELPAPNRCATSEECYCGE 4428
QY 2908 TVVD-GNDATCKALHLSKIVNIPSSQVLRGSSNIITQATVFNAAALVLSRESKDVVFG 2966
Db 4429 KVISFSKELTKITRIAK-----QHVVITNTVLQGIWGMILAKYKNTDEVFG 4476
QY 2967 RYVSGRQLPVE-YQDIVGCTNAPVPRAHIES-DYNQLLHDIQOYLLSLPHETIGFS 3024
Db 4477 TVVSGREA-PVQDIEEMVGLFTIIPTRISFEGARSFKVLEKKTQAEIESNESYSYNNLS 4535
QY 3025 DLKRNCTDPEATNESCCTIYNFEYHPESQEQVEMGLVTKFVNTEMDEPLVDLAI 3084
Db 4536 EIOVLSEMKRELTH-----VMAFONYAF-DEELFRSQSGETGFELEGVH-GKERTNTNENL 4590
QY 3085 AGEVPEPDGAGLKV-----TVIAKTQLFGRKRVEHLEEVSKTFEGNL 3126
Db 4591 TGVLEDEQLKLTFTENYVDNTIETLEKHIITVAEQVADETTQTLRDIN 4641

RESULT 4

TYCB_BACBR ID TYCB_BACBR STANDARD; PRT; 3587 AA.
AC G30408;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrocidine synthetase II [Includes: ATP-dependent proline adenylase
DE (ProA) (proline activase); ATP-dependent phenylalanine adenylase
DE (PheA) (phenylalanine activase); ATP-dependent D-phenylalanine
DE adenylase (D-PheA) (D-phenylalanine activase); ATP-dependent racemase
DE [ATP-hydrolyzing] (EC 5.1.1.11)].
GN TYCB.
OS Bacillus brevis.
OS Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
OX NCBI_TaxID=1393;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 8185;
RX MEDLINE=98012987; PubMed=9352938;
RA Mootz H.D., Marahiel M.A.;
RT "The tyrocidine biosynthesis operon of Bacillus brevis: Complete
RT nucleotide sequence and biochemical characterization of functional
RT internal adenylation domains.";
RL J. Bacteriol. 179:6843-6850(1997).
CC -!- FUNCTION: ACTIVATES THE SECOND TO FOURTH AMINO ACIDS IN TYROCIDINE
CC (IN TYROCIDINE A, PRO, PHE, AND D-PHE) AND EPIMERIZES THE LAST
CC ONE.
CC -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine = AMP + diphosphate + D-
CC phenylalanine.
CC -!- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES (BY
CC SIMILARITY).
CC -!- PATHWAY: Cyclic peptide antibiotic tyrocidine biosynthesis.
CC -!- SUBUNIT: LARGE MULTIZYME COMPLEX OF TYCA, TYCB AND TYCC.
CC -!- DOMAIN: CONSISTS OF THREE MODULES, INCLUDING AN C-TERMINAL
CC EPIMERIZATION DOMAIN. EACH MODULE INCORPORATES ONE AMINO ACID INTO
CC THE PEPTIDE PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS
CC RESPONSIBLE FOR SUBSTRATE ADENYLATION, THIOLATION, CONDENSATION
CC (NOT FOR THE INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND
CC N METHYLATION (OPTIONAL).
CC -!- MISCELLANEOUS: TYROCIDINE IS A MIXTURE OF FOUR CYCLIC
CC DECAPEPTIDES, TYROCIDINE A (D-PHE-PRO-PHE-D-PHE-ASN-GLN-TYR-VAL-
CC ORN-LEU), B, C, AND D, IN WHICH PHE, AT POSITIONS 3, 4, AND TYR
CC RESIDUES ARE GRADUALLY REPLACED BY TRP, DEPENDING ON THE RELATIVE
CC CONCENTRATIONS OF THESE AMINO ACIDS IN THE GROWTH MEDIUM.
CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AF004835; AAC45929.1; -
CC HSSP; P14687; IAMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00501; AMP-binding; 3.
DR Pfam; PF00550; pp-binding; 3.
DR Pfam; PF00668; Condensation; 4.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR PROSITE; PS00455; AMP-BINDING; 3.
DR PROSITE; PS00075; ACP_DOMAIN; 3.
KW Ligase; Isomerase; Antibiotic biosynthesis; Phosphopantetheine;
KW Multifunctional enzyme; Repeat.

772	PGQAIPELPPIGRPIISNTGVIILDEG--LQKPEGIVGELYISCANVGRGLYHQPELTAE	829
842	PFETDIPSWYPANTFPDGAKYRTGDLARVASDGSIVCLGRIDISOVKIRGORVELGAJET	901
830	KFLD-----NPYQGERMYRTGDUALMDPGQLEFLGRIDHQWKIRGRIELGIES	881
902	HLRQOMPDDLTIVVEATKRSQSANSTSLIAFLIGSSYFGNRPDSAHILHDKATKAINIKL	961
882	RLLNHPAIKEAVVID--RADETGGKFLCAVVLQKALSDEMRAY-----L	925
962	EQVLPRHISPFYICMLELPRTAGTKIDRRRLRMKGDKIDLKQTOGAIVQOAPAPIPVFA	1021
926	AQALPEYMPISFFVLEPIETVPTNGKTDRAAL-----PKPEGSAKTADYVAPT-T	975
1022	DTAAKLHSIWQSLGIDPATVNVGATFELGGNSITAIKMVN-MARSVGMDLKVSNYIOH	1080
976	ELEGKLVAIWEQILGVSP--IGTDHEFTTGGHSLKAIQLISRIQKECQADVPURVLFQ	1033
1081	PTLAGISAVYK-GDPLSYTLPIPKSTHBEQVQESYSGQRMLFPLDQLDYGSWMYLIPIYAVRM	1139
1034	PTICALAAAYVEGGEESAYLAIPQAEPOAYVPVSSAQKRMLILNQLDPSHTVYNLPVAMIL	1093
1140	RGPVNDALRRALAALEQRHETUETRTFEDQDGVGVQIVHBEK-----LSEEMKVYIDLCL	1191
1094	EGTLDKARLEHAISNLVARHESURTSFHTINGEVPVSRHBEQGHLPVLYLETAEEQVNEVI	1153
1192	GSDLDPEVLNOBOTTPFNLSSEAGWRATILLRGEDDHLITVMVHHIISDGSWIDLVRD	1251
1154	LGFMPQPDLV---TAPL-----CRGLVUKLAENRHVLIDMHIIISDGVSSQLIIND	1202
1252	LNOLYSAAKSDKDPALSALPTPIQYSDFAKWOK--DQFIEQEQLQWYKKQLKSDSPA-	1308
1203	FSRLYQN-----KALPEQRHYHKDFAVWEKAWTQTTDYQKQEKYWLDRFAGEIPVL	1253
1309	KIPTDFARPALLSGDAGCVHVTIDGELYQSIRAPFCNEHNTTSFVLLAAAPRAAHRYUTAV	1368
1254	NLPMDYPRPAVOFEGERYLFETKEQLLESLODYAOKTGTTLVMVLLAAVHLLSKYSQG	1313
1369	EDAVIGTPIANRNRPELIEDIIGCFVNTQCRINIDHDHDTGTLINOVKATTTAAFNEDI	1428
1314	DDVMIGTVTAGRVHPDTFESTMGFMVNTLAWNQSAPTFTFQFLELVKDNTLAFAFEGQY	1373
1429	PFERVYSALOPGSRDLSSTPLAOLIFAVHSOKDLGRFKFQGLSEVP-VPSKATYTRDMEF	1487
1374	PFEEELVEKLAI-QRNSRNPFLDTFLTLQNN-DADLIELDGLTVPVYVEGEVAKFDLSL	1431
1488	HLFQETDSLKGSVNFADLEFKMETVENVRVFEILRNGLOSSRTPVSIPLDTGIVTLE	1547
1432	EASENQAGLSFCFECTCKFARETIERMSLHYLIQILO--AVSANTEQELAQI-EMLTAFE	1488
1548	KLDVL---NVKHVDYIPRESSLADVFQTSAYPDSLAVVSDSSCLTYTLEDROSDILAGW	1604
1489	KQELLVHFNDAALYPAESTLSOLFEDQAOKTPEQTAVVFGDKRLTVRELNERANQLAHT	1548
1605	LRRRSMPAETLVAVFAPRSCETTIVAFPGVLKANLAYLPDVRSPSARVQDILSGLSGPTI	1664
1549	LRAKGVAEQSVGIMAQRSLEMAIGIITALKAGGAYVPIDPDYPNERIAWLEDCR--RL	1606
1665	VLICHTDAPDPIEVTVNVEFVRIIRDALNDSNADGFEVIEHDSKTP--SATSLAYVLYTSGS	1722
1607	VLTTQOOLAA--EKMTAVNECYLEL-----DEEGSYSPQTEINPIHTAADLAYIITYSGT	1657
1723	TGRPKGVMIEHRVIRIIVT-----SCIPNYPSETMAHMAHTAFDGAASYEY	1770
1658	TGRPKGVMVZHRGIVNSVTNNRDEFALSVRDSGTL-----SLSAFONAFALTF	1706
1771	SALLFGRTLVCVDMYMTTLARALKDVFREFHVNAAASHVTSSQDVPRLVRPRRLSRTLMFF	1830
1707	TLIVSGSVVLMPDHEAKDFIALRNLIAWECYSVYVFPV-----MF	1748
1831	FLVVTDSTAPDALDAQLYOG-----VQCYNGYGPTENGWSTIYPI	1872
1749	QAILECSTPADIRSIQAVMLGGEKLSPLKVLQCKAMHPQMSVMNAYGPTESSYMAT--YLR	1807

1873 DSTESFNGVPIGRALNNSGAYVVDPEQQLVIGVGMGELVVTGDLARGYSDK-ALDENR 1931
1808 DTQPD--QPTIGRPIANTAIYIVDQHHQLLPVGVGEICIGGHLARGYWKPELTAEK 1865
1932 FVHTVNDQT--VKAYTGRVRYRIGDGLIEFGMDTQFKRGNRIESAEATEAALRD 1989
1866 FV---ANPVPGERMYKTGDLGRW-LHDGTIDFGRVDDQIKRGYRIEVEGEIEAVLLY 1921
1990 SSVRDAAVLQONE----- 2003
1922 DQINEAIVVAYQDGRGDSYLAAYVTGKTAIESESLRAHLRLRELPAYMVPYLIQLDAFPL 1981
2004 -----DOAPEILGFVVADH----- 2017
1982 TPNGKVDKALPKPEGPATGAAYVAPATEVEAKLVAIWENALGICSGVGLDHFELGGH 2041
2018 -----DHSENDKGQS----- 2027
2042 SLKAMTVVAOVHREFQIDLLKOFFAAPTTRDLARLIEHSEQAAGAAIQPAEPQAYPVS 2101
2028 -----ANQVEGWODHFES-GMYSDIGEIDPSTIGSDFKGWTSYMDGSOIDFDEMHEW 2078
2102 SAQORWYLLHLEGAGISYNTPGIIMLEGKLDREQLANALQALVDRHDLILRTSEWVGD- 2160
2079 LGETTRTLHDNRSLGNVLEIGTSGMILFNLDLSRLESYVGLPSR-----SAAAFVNKATES 2135
2161 --ELVQKIHDR-----VAVNME-----YVTAEEQQIDDLFHFVRFPLDS 2198
2136 IPSLAGKAKYQVG-----TATDIGOVDDLHPLDVLV-----NSVTOY----- 2172
2199 VPPLLRLMSVLKADERHLLYDMHHIAADAASITILFDELAELYQGRELPDEMRIQKQDFA 2258
2173 -----FPSSE--YLAETADTLHL-----PNVQRIFFGDVRQATNE----- 2207
2259 VWQKALHESDAFKQEAAYWLSTFAGNITAVDFTDPRPAVKFAGQVTLSDMQELLSA 2318
2208 -HFLAARAIHT-----LGNATKDDV-----RQMAE----- 2233
2319 LHLEAAHTNTLFWLLAAYNVLLAKYAGQDDIIVGTPISGRSKRAELAPVGVFVHTLAI 2378
2234 -----LEDWEEELLVEPAF-----FTSLKD-----RPPG-----LVE 2260
2379 RNKPTAEKTFQLEQVKQNAL--DAFDHQDYFESLVEKLGIPRODGRNPLFDTMFIQ 2436
2261 HVEILPKNM-----EAVNELSAYRYAAVHVHVGSLGD----- 2292
2437 NDELHAKTLQLVYRPVSDSALDVAKFDLSFHLTERETDLFLRLEYCTKLFKQQTVERM 2496
2293 -----ELVLPVEK-----DDWIDFQANLQNO-----KSLGDL 2319
2497 AHFLQILRAVTANPENLEQIEMLTAAEQMILLVAFNPDTHREYRADQTTIQQLFEELAEK 2556
2320 LKSSDAAIMAVSKIPF-----EITAPER-----QVVASLNSIDEMQLSTIRSSA 2364
2557 MPHTALVPEEKMSFRELNERANQALAVLRKGVGPAQIVALLVRSABMWATLATLAK 2616
2365 EGDSSLSV-PDI--FRIAGEAGFRVSSARQWSQNGALDAVPHHCCSQGRTL----- 2414
2617 AGGAFLPVDPDYPERIR-----YMLEDSQAKLVVTHAHL---LHKVSSQSEVVVDVDDPGS 2669
2415 -----VNPTD-----HHL----- 2423
2670 YATQTNLPCANTPSDLAYIITGTTGPKGVMLHKGVANLQAVFAHLGVTPQDRAG 2729
2424 -----RGSDLLTNRPLQRLN-----BRIAEVRLERSLL-----PSYMI-- 2459
2730 HFASISFDSAVMDFGPLLSGATLYVLSRDVINDFORFAEYVRDANITELTLPPTAIYL 2789
2460 -----PSNIVLDKMP-----LNANGKYDRKELSRRAKVVYPKQOATAAPL 2498
2790 EPEQVPSLRLITAGASSVALVDKWEKVTYVNGYGPTSTVCATLWAKAPDE-----PV 2845

2499 PTF-----PISEVEVILCEEATEV-----FGMKVDIITDHF-----NL 2531
2846 ETITIGRPIQNTKLYIVDDQLQKAGOMGELCISGLARGYWNRELTAEKFDVPFV 2905
2532 GGHSLAT-----KLISRIDQRLKVR-----ITVKDVF----- 2559
2906 PGTKMYRTGDLARWLDPDGTIEYLGRIDHQVKIRHRYELGEVESVLLRYDTVKEAAALTH 2965
2560 --DHPVFADLASVIRQSLG-----QOPVSDGQOGDKSA- 2591
2966 EDRGQAYLCAYVAEGEATPAQLRAYMENELPNMVPAPAFIQLEKMLPTPNDKIDKAL 3025
2592 -----HMAPRTETAILCDEFKVLGF-QVGITDNFFDLGGHSLMATKLAVRI 2638
3026 PKPQNEENRTEQYAAQPOTELEQLLGIWADVIGIKVGTQDNFFELGDSIKAIQVSTRL 3085
2639 GHRLDITVSVKDFDHPVLQALALDNLVQSKTNEIVGGREMAEYSPFQLLFTEDPEEF 2698
3086 -NAGWMTLAMELKFQYPTIEAALRV-----IPNSRE----- 3116
2699 MASEIKPQLELQELIQQIDYIPSTOMOKAFLFDHTTAR---PRPFVFPVID-FPSTSEPDAA 2754
3117 --SE-----QGVVEGEIALTPIQKWFANFTDRHHWQAVMLFREDGFDE----- 3160
2755 GLIK-ACESLVNHLIDFRTVFAEASGELYQVWLSCLDLPQIVTETEDNINTATNEF----- 2809
3161 GLVROAFQOQIVEHHDALRMVYKQEDGAIKQINRGLTDERFR-FYSYDLKNHANSEARILE 3219
2810 LDEFKPEPVLGH-PLIRFTIITKQKSMRVMIRSHALYDGLSLEHVVRKL-----HMLY 2863
3220 LSQIOSSIDLEHGLVHVVALFATKGDHLLVAIHLLVVDGVSWMRILFEDFSSAYSQALH 3279
2864 NGRSLAPP-----HQSRYMQYTADGRE--SGHGFWRDVTQNTPTMTILSDTVVDGNDAT 2916
3280 QQEIVLPKKTDSFKDWAQKQKYADSDELLREVAYWHNLETTTITTAALPTDFVT---AD 3335
2917 CKALH---LSKIVNPSQVLRGSSNIITQA-----TVFN-----AACALVLSRESKDVV 2964
3336 RKQKHTRTLSFALTVPQ-----TENLLRHVHHAYHTEMNDLLLTALGLAVKDWHTNGVV 3390
2965 FGRIVSGROCLPVEYQ-DIVGCTNAVPVRAHTESSDYNQLLDIOD--OYLLSLPHET 3020
3391 INLEHGREDIQENMNVRTIGWFTSOYPVVLDMERAE--DLPYQIKOTKENLRRIKPKG 3448
3021 IGFSDLKRNCWDN--PEATINFSCCITYHNFYHPESQFQOQVEMGVLTKEFVNIEWDE 3077
3449 IGVEILRTLTSQLOPPLA-----FTLRPEISFN----- 3477
3078 PLYDLAIAGEVDPG--AGLKVTVIAKTQFLGKRKRVEHLLLEEVSKTPE 3123
3478 -----YLGQFESDGTGGTFTSPGLGTQGLFSPESERVFLDISAMIE 3519

RESULT 5

PPS3_BACSU

ID PPS3_BACSU STANDARD; PRG: 2555 AA.
AC P39847;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Peptide synthetase 3.
GN PFSC OR PPS3.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
Borriass R., Boursier L., Brans A., Braun M., Brignon S.C., Bron S.,
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,

[illegible]

RL Mol. Microbiol. 8:821-831(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RC MEDLINE=97124189; PubMed=8969502;
RA Yamane K., Kumano M., Kurita K.,
RT "The 25 degrees-36 degrees region of the Bacillus subtilis
RT chromosome: determination of the sequence of a 146 kb segment and
RT identification of 113 genes."
RL Microbiology 142:3047-3056(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogawara N., Moser I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertoletto M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denton K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scantlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Solido B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis".
RL Nature 390:249-256(1997).
RN [5]
RP SEQUENCE OF 514-800 FROM N.A.
RC STRAIN=ATCC 21332;
RX MEDLINE=92290255; PubMed=1601288;
RA Borchert S., Fatil S.S., Marahiel M.A.;
RT "Identification of putative multifunctional peptide synthetase genes
RT using highly conserved oligonucleotide sequences derived from known
RT synthetases."
RL FEMS Microbiol. Lett. 71:175-180(1992).
CC -1- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME ABLE TO
CC ACTIVATE AND POLYMERIZE THE AMINO ACIDS LEU, GLU, ASP AND VAL.
CC
CC -1- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHATETHEINES.
CC
CC -1- PATHWAY: Cyclic peptide antibiotic surfactin biosynthesis.
CC
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY
CC
CC -1- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; D13262; BAA02523.1; -
DR EMBL; X70356; CA449817.1; -

Query Match

11.18; Score 1793.5; DB 1; Length 3587;

Best Local Similarity

22.4%; Pred. No. 2.8e-95;

Matches 732; Conservative 458;

Mismatches 1130; Indels 953; Gaps 99;

Db 1132 NVELOIAYSE--STEDOVERIAIEFMOPFALEVRPLLRVGLVLEAEERHLFIMDMHHII 1188
QY 1240 SDGWSIDVLRDLNOLYSAALKSKDPLSALPLTOYSDFAKW-----OKDOFTEOEKQ 1294
Db 1189 SDGWSQWIMQIADLYE-----KELPTLGQYKDFVWHNRLLOSVDIEKQAS 1239
QY 1295 L--NYWKQKOLDSPAKIPTDFARPALLSGDAGCVHVTIDGELYQSRAFNCNEHNTTSFV 1352
Db 1240 LAERICRDSSTESTDRLPKYOPFKALMVKDLHVSQESSLWMIYTRWQOKQOQHYIWFYL 1299
QY 1353 VLLAAFRAAHRLTAVEDAVICTPIANRNRPELEICGFYNTQCMRINIDHDTFGTLI 1412
Db 1300 LRIMEFELS---KYGODDIVGTPIAGRSHADVENMLGMFVNTLAIRSLNNEDETFKDFL 1356
QY 1413 NOVKAATTAAFENEDIPPERVVSALQPG--SRDLSLPLAQLLFAVHSOKDLGRFRFOGLE 1471
Db 1357 ANVKOTALHAYENPDYPTDLVEKL--GIQRDLSRNPFLDFMFWLQON-TDRKSFEVEQIT 1413
QY 1472 SVP-VPSKAYTRFMEHFLFQETDSKGSVNFADLFKMETVENVVRVFFELIRNGLOSS 1530
Db 1414 ITPYVPNRSRHSKFDLTLEVSERQNETLLCLEYCTKLTDTKTVERMAGHFLQILHAIVGN- 1472
QY 1531 RPPVSILPLTDCIVTLEKLDVL--NVKHVDYPRESSLADVPOTQVSAYPDSLAVVDSSC 1587
Db 1473 --PTIIEIIEELSEEEKOHLIFEENDTKTTPYLPCKQKDYLRNRRRRADHVAVWGKQD 1530
QY 1588 RLTYTELDRQSDILAGWLRRRRSMPAETLVAVPAPRSCETIVAFEGVLKANLAYLPDVR 1647
Db 1531 TLITYRELNERANQAVRLKQGVQDNIVGLLVERSPEMLVIMGILKAGGAYLPDPEY 1590
QY 1648 PSARVODILSGSGPTIYLIGHDTAPPDIEVTNVEF---VRIRDALNSNADGFEVIEHD 1704
Db 1591 PADRISYMQD--CGVRIML---TQOHLLSLVHDEFDCVILDEDSLXKGSNLAAPVN-- 1643
QY 1705 STKPSATSLAYVLYTSGTGRPKGYMIEHRVIRIIVTSGCIPNYPSETRMAHMAIADG 1764
Db 1644 ---OAGDLAYIMYTSSTGPKGYMIEHRVIRIIVTSGCIPNYPSETRMAHMAIADG 1699
QY 1765 ASYEIYSALLFQRTLCVDYMTTLARALKKOVFFREHVNAAHSVTSOSSODVPLRVRRLS 1824
Db 1700 LTFEFGSLHGAELPYTKVLLDAEKLHREL-----QANQIIMWLTSPLF--NOLS 1751
QY 1825 RFLMFF-----LVVTSTADPAL--DAQLYQGVQCYNGYNGYPTGVNSTIYPIDSTES 1877
Db 1752 QCTEEMFAGRLSLIVGGDALSFKHNNVKKCPNLTMMNGYPTENTTFSTCLFD--KE 1809
QY 1878 FINGVPIGRALNNGAYVVDPPQQLVGIGVGMELVVTGDLARGYSDK--ALDENRFVHT 1936
Db 1810 YDDNIPKAKANSVYIMDRYQQLQPVGVGELCVGGDVGARGYMNQOPALTEERFVPNP 1869
QY 1937 VNDQTVKAYRTGDRVYRIGDGLIEFFGRMDTQFKIRGNRIEASIEAALLRDSVVRDAA 1996
Db 1870 PAPGE--RMVRTGDLARW-LPDGTIEYLGRIQOVQKIRGYRIEPEGEITLLVKKHKKVESV 1927
QY 1997 VYLOQNEQAAPILGFVADHD-----HSENDK-- 2024
Db 1928 IMVVEDNNGQKALCAIYYPPEEVTVSELREYIAKELPVMYPAYFVQIEQMLTQNGKVN 1987
QY 2025 -----GOSANQVE-----CQODHF----- 2038
Db 1988 RSALPKPDGEFTATYEVAPSSDIEMKLAIEIWHNVGLVKNKIGVLDNFFELGHSLRAMTM 2047
QY 2039 -----ESGMDSIDGIEIDPST----- 2053
Db 2048 ISOVHKEFDVELPLKVLFTPTISALAOYIADGQKMYLAIOPTVPTDYYPVSSAQKMY 2107
QY 2054 IGSDPKGWTSMYDGSQIDPDE-----MHEWL-----GETTRTLHD 2088
Db 2108 ILYEPEGAGITVNPVNMVFIIEGKLDYQRFEYAKSLVNRHEALRTSFYSLNAGEPQVRVHQ 2167
QY 2089 NRSLL-----GNVLEIG-----TGSCMILFNLSR----- 2112
Db 2168 NVELOIAYSEAKEDEIEQIVESFVQFDFLEIAPLLRVLGLVKLASDRYLFELMDHHIISDG 2227
QY 2113 -----LESYVGLEPS-----RGAAPVFNKATESIPSL 2139
Db 2228 VSMQIITKEIADLYKGKELAEHLIOYKDFAVWQNEWFQSDALEKQKTYWLTFAEDIPVL 2287
QY 2140 -----AGK 2142
Db 2288 NLSTDYPRPTIQSFEGDIVTFESAGKQLABELKLAETGTTLYMLLLAAYNVLLHKSQ 2347
QY 2143 AKVQVGT-----TDIGQVDDLHPD-----LVVLSVTOYPPSSEY 2178
Db 2348 EIVVGTPTIAGRSHADVENIGVMFVNTLAKKTPIAVRTFHEFLLEVKONALEAFENODY 2407
QY 2179 LAE-----IADTLHLPNV-ORIEFG-----D 2199
Db 2408 PFENLIEKLOVRDLRSRNPFLDTMFSLSDNIDQVEIGIEGLNFSPEYMOYWTAKFDISFD 2467
QY 2200 VRSQATNEHF-----LAARAHTLTKGNATKDDVRQKMAEL-----EDMBE 2239
Db 2468 ILEKODDIQOFYENYCTNLFKKTETIERLATHEMHILQEI VINPEI--KLCEINMLSEEOQ 2525
QY 2240 ELLVE-----PAFETSILKORFPGLVEHVEILPKNMEAVNELSARY----- 2280
Db 2526 RVLYDFNGTDTATYATNKIFHELFEQVEKTPDHIADVIDEREKLSYQELNAKANQALRYLR 2585
QY 2281 -----AAVHV 2286
Db 2586 QKGVOPNSWGMVDRSMDIMVGMGLVKAGGAYPIDIDYDQERISYMWEDSGALLIT 2645
QY 2287 RGLSDEL-----VLPVEKDDWIDFOANQLN----- 2312
Db 2646 QOKLTQOIAFSGDILYLDQEEMLHEEASNLPIARPHYIAYIYTSGTGKPKGVMIHQ 2705
QY 2313 -----QKSLGDL-----LKSSD 2324
Db 2706 SYVNAMAMKDAYRLDTPFVRLLQMASFAFAFDVSAGDFARALLTGGQLIVCPNEYKMDP 2765
QY 2325 AAIMAVSKIPFEITAFER--QVVASLNSNIDWQL--STIRSAEDSSLSVDPIDFRIAG 2380
Db 2766 ASLYALIK-KYDITFEATPALVPLMEYIYEQKLDISQLOLILVGSDCSMEDFKTLVS 2824
QY 2381 EAGFRVSSAFQWSONGALDAVHHCCSQGRTLNVFP--TDHHLRGSDDLTHRPLORLON 2439
Db 2825 REGSTIRIVNSYGVTEAALILAI-----MNQPLSLRHVTGT--VPICKPYANMK- 2871
QY 2440 RRIATEVERLRLSLPSYMPISNVIVLDKMLN---ANGKVDREKLSRAKVPK--- 2491
Db 2872 -----MYINQYLIQIPVGVIGELCIGGAGVARGYLNRPDLTAE-KFVPPFPV 2919
QY 2492 -----QOTAPLPTFPISEVEVI-LCEEATEVEFCMKVDITDHFNFNGLGH-SLLATKLISRI 2545
Db 2920 GEKLYRTGLARMPDGNVFEFLGRNDHQVKIRGIRTELGEIEAQLRKHDSIKEATVIARE 2979
QY 2546 DORLKVRITVKDFVHPV-----FADLASVTRQSLG-----LOQ 2579
Db 2980 DHMKYKLCAYMTEGEVNVNABELRAYLANDRAAMIPSYEVSLEAMPLTANGHIDKRLPE 3039
QY 2580 P---VSDGOGQDSAHMARTETEAITLCEDEFAKVLGFQ--VGITDNNFDLGGHSLMATKLA 2635
Db 3040 PGCSISIGTEYDR-----PRTMLEKLEBIWKDVLGLQVRGILHDDFFTIGGHSKAMAYI 3094
QY 2636 VRIGRLDPTSVKQVDFHPVLFQALALDNLVQSKNTNEIVGGRMAEYSPFQLLFTEDP 2695
Db 3095 SQVHEKQCTEVLRLVLFETPTIQLAKYIE-----TDT 3128
QY 2696 EEFMASEIKPQLELEQII--QDIYPTOMOKAFELDHTTARPPFPVFFYIDF--PS---T 2748
Db 3129 EOYMA-----IQPVSGQDYYPVSSAQKRFI-----VNOQFVGISYNNPMSIMLJE 3174
QY 2749 SEPDAAGLAKACESLVNHLDIERTVFAEASGELYQVVLSCDLPQIVIEDININ---T 2804
Db 3175 GKLERTRUESAFKRIERHERSRTSFEIINGK-----PVOKTHEEVDNMSYOV 3223

Qy	474	RABIESNWSQPLEVQDTHLH---HEML-----KAVSH-----SPTKTAIQWGDWT	517
Db	500	-----BEKDIIILTFENHEKTDGPKNTLSRLFEERAETKPDHTAVIFEDQOLT	547
Qy	518	YSELNDVSSRLAVHIIKSLGRUAQAAIPIVYFEKSKVVIASMLAVLSGNAFTLIDPNPP	577
Db	548	YRELNEKANOLALLREKGVK-PDTIVAIMTDRSLEMIIGIILKAGGAYLIDPDYPE	606
Qy	578	ARTAOVOTPRATVALTSKLHRETVOKLGRVGVVDDDELQSVASDDFSSLPKSDOLAY	637
Db	607	DRVKYMLEDSGADWVIOEFPK---SKIDGRQUITAED-TRSPS-KENLPNVNKASDLAY	661
Qy	638	VIFTSSTGDPKIMIEHRAFSSCALKGASLNGSDTRALQFTHAFGACCLIEIMTTLI	697
Db	662	VIVTSGSGRPKGVMTTHRNWVHYVDAFTKRIPLSEHDTVLQVVSFSDAFSEEVYPILA	721
Qy	698	NGCVCIPS--DDDRMNSIPSFNIRNVNMMATPSYMG-----TFSPEDVPGUATL	747
Db	722	CSGRLVISRKVSJLNDLDELVTICKYKRVTLVSCPLLNIDRKNQHLTFHPQMKFTSGGD	781
Qy	748	VLVEOMSSSVNAIAPKQLLNGYGOSSESSICFASNMSTPNP-----MGRAVGAHSW	802
Db	782	VLKFEVENIIGK---ADVNSYGPTEATVCATYYQLSSADRKKTSIPIGAPLSNYK-	835
Qy	803	VIDPNDINRLVPIGAVGELVIESPGIARDYIVPPPEKSPFTDIPSWYPANTFPDCAKL	862
Db	836	VYIADQVGRPQVPGVPGELLIGCEGVARGVNLHETLTAKAAVVD-----ESGERV	885
Qy	863	YRTGDLARVADSGSIVCLGRIDSOVKIRGORVELGAETHLROMQDDTLIVVEATKRSQ	922
Db	886	YRTGDLARVADSGSIVCLGRIDSOVKIRGORVELGAETHLROMQDDTLIVVEATKRSQ	943
Qy	923	SANSTSLIAFLGSSVFGNRPSPDAHILHDHATKAINIKLEOVLPRHSIPSYICMLELPR	982
Db	944	E-NSKYLCAVIA---FNNKAD-----IQOVBRLAKDLPEYMIPSCFINKLDOIPR	990
Qy	983	TATGKIDRRRLRMKIDLDKOTOGAIVQOAPAPIPVFADTAALKHSIWQSLGIDPA-T	1041
Db	991	TINGKADLKAL-----PEPDRAFAQARYEAPR-----NQTEALLLSIWQDIL--PAEQ	1037
Qy	1042	VNVGATFFELGGNSITAIKVMNARS-VGMDLKVSNYIQQHTLAGISAV-----VKGD	1093
Db	1038	IGINDHFFDIGHGSLKAFSAKIAQSAKVEVTLKEIFNHSHTODLAAIYIAQKOKQVSD	1097
Qy	1094	PLSYTLIPKSTHEGPEQSYSGRLMFLDQDVGSLWYLIPIYAVRMGRPVNDALRRALA	1153
Db	1098	-----LOKAEKEYPLSSAOKRLYLINOIEGOTAYNPPFAMKINGELQTDKAEKFR	1151
Qy	1154	ALBQRHETLRTTFEDQDGVQIVHEKLSSEMKVIDLCGSDLDPPFVNLNOEQTTFFNLSS	1213
Db	1152	TLIKRHESSRTSFVTINGEPVQVINEEVTPEMKYRELDNCSLR-ERMNQ-FIRPPELEK	1208
Qy	1214	EAGWRATLLRLGEDDHILLTVMHHIISDGSIDVLRDLNQLYSAAALKSDKPLSALTPL	1273
Db	1209	APLLRAELVRVNAEHLDDMHHSIDGVSIGILMKEW-----AALYEKE---LAPL	1259
Qy	1274	PIQSDPAKWKQDOFTBO--EKOLNWKYKQKLDSSPA-KIPTDFARPALLSGDAGCVHVT	1330
Db	1260	KIQKYSEWQRPQWQDKLKEESWLSVFQNDIPVLNMPPTDFPRQMSYEGDRIAFIA	1319
Qy	1331	IDGLOSRAFNCNEHTTSFVVLAAFAAHRYLTAVEDAVIGTPTIANRNPREDIIG	1390
Db	1320	IERELTDKLTAKENGVTWMLLAGYTTILLSKYTGQEDIIVGSPAGTRTREETQVTG	1379
Qy	1391	CFVNTQCMRINIDHDTFTGLINQVATTAATTAENEDIPERVVSAALQPSRDLSSPLA	1450
Db	1380	MFVGTAMRNHPKGGRTFIEYLQDVKENTFNAYENQDYPDELVDKLDL-ERDISRNALF	1438
Qy	1451	OLIFAVHSOKDLGRFQGLSVFVPSK-AYTRFDMEFHLFOETDSLKGSVNADELFKM	1509
Db	1439	DTWFDQALDD-APDIEGLUHVPEVDFLEFOISKFDLSLTAESAGVITTFHLEFCTRLYKK	1497
Qy	1510	ETVENVVRVFEILR---NGLQSRSTPVSILPLTDGIVTLEKLDVLVNKHVDVPRESSLA	1566
Db	1498	ETAETLQAHEVNILROISDHPQKTLNDISMLSEERHTVLYQFNDTNEH-----PSGIFS	1553
Qy	1567	DVFTQVSAYPDSLAVVSDSRLTYITELDRQSDILAGLRRRSPATLVAVPARSCET	1626
Db	1554	ELFEEQAEKSPNHPAAVFKDQMLTYRELNEKANQARTLRQGVQRESVVGIMAEARSLEM	1613
Qy	1627	IVAFEGVLKANLAYPLDVRSPSARVODILSGSGPTIVLIGHDTPDPIEVTNVEFVRI	1686
Db	1614	LTGILAVLKAGGAYMPIDPGLPKERIOYLTD-SGADLLLTQHOL-----IGSIFA--	1664
Qy	1687	RDALNDSNADGFEV---IEHDSKTSATSALAYLYTSGTGRPKGYMIEHRVIRTVTS	1742
Db	1665	GEIIQIDOADAYDTDGSNLEHLN---SPGDLAYVIYTSGTGNPKGVMEHRNIHA---	1718
Qy	1743	GCIPNY-----PSETRMAHWATIAFGASVEIYALLFGRTLVCVDYMTLIDARAL	1793
Db	1719	----HYTWKRYHELASFVNLLQLASMSFDVAFGDLCSRLNGCTMYIVDDVKLEMLNL	1774
Qy	1794	KDVEFRHVNAASHVTSQDQVPLRVPRRLSRTLMEFF---LVVTDSTAPDALDAOGLY-	1849
Db	1775	YDMINKYI---HMLESTSLIPLMKYIDHHKLDFFSMKLLINGSDTCTIKDYKWLVE	1830
Qy	1850	---QGVQYNGYGTENGVMSTIY--PIDSTESFINGVPIGRALNNSGAYVDPDQOLVG	1904
Db	1831	REGORMRIINSYGVTEASVDSGYVEEALDRIPETIAN-TPICKPLDNTAFYILDPSLNQ	1889
Qy	1905	IGVWGLVTVTGDGLARGYSK-ALDENRFVHITVNDQTV--KAYRTGDRVRYRIGDGLIE	1961
Db	1890	GVYGYELVIGEGIARGYLNKPELTKERFV---PNRFAAGNMKYTGDLARW-LPDGNVE	1945
Qy	1962	FEGRMDTQFKIRGNRIESAELEALLRDSVRAVVAVVAVVAVVAVVAVVAVVAVVAV	2021
Db	1946	FLGRIDHQAIRGRIETGETEYKLEENQISEAVVI-----DRE	1985
Qy	2022	NDKQSANQVEGQDHFESGMYSDIGEIDPSTIGSDFKGMTMYDGSQIDFDEMHEWLGE	2081
Db	1986	DKKG-----HKY---	1992
Qy	2082	TTRTLHDNRSLGNVLEIGTSGMILFNLDLSRLESYVGLSPSRAAFVNKATESIPSLAG	2141
Db	1993	-----	1992
Qy	2142	KAKVQVGTATDIGOVDDLHPDLVVLNSVIOVFPSPSEYLAETADTLIHLPNVQRIFFGDVR	2201
Db	1993	-----	1992
Qy	2202	SQATNEHFLAARAIHTLGNKATKDDVRQKMAELEDMEELVEPAFTSLKDRPGLVEH	2261
Db	1993	-----	1992
Qy	2262	VEILPKMNEAVNELSAYAAVHVHVRGSLGDELVLPVEKDDWIDFQANQLNOKSLGDLK	2321
Db	1993	-----LCAY---	1996
Qy	2322	SSDAAIMAVSKIPPEITAFERQVAVVLSNIDENQWLSIRSSAEGSSLSVDFIARIAGE	2381
Db	1997	-----IVARAKTN-----	2004
Qy	2382	AGFRVEVSSARQWSQNGALDAVHFHCCSQGRTLVNFPTDHLHRLGSDLLTNRLQRLQNR	2441
Db	2005	-----TN-----	2006
Qy	2442	IATEVRELRSLPSYMPISNIVLDKMLNANGKVDKELSRRAKVVVPKQOQTAAPLPTF	2501
Db	2007	---ELREYLDHCLPDYMLPSYFTQINKMPLTPNGKIDOKALP-----	2045
Qy	2502	PISEVEVILCEEAEVFGMKVDTITDHFENLGGHSLATKLISRIDQRLKVRITVKOVFDH	2561
Db	2046	-----EPAGDVIA-----	2053
Qy	2562	PVFADLASVIRQGLGLOQPVSDGQGDRAHMAPRTETAILCDEFKVLG-FQVGTIDN	2620

Db 2054 -----ASGVEAPRNETEEKLAADVQVFLDRDKIGINDN 2086
 Qy 2621 FFDLGGHSLMATKLAVRIGHRLDTTVSKVDHDFVLQFALADLNLVQSTNETVGGRE 2680
 Db 2087 FFEIGGDSIKALQIVSKLS-RADLKLVQKDLFTNPFIRHLSKYVKETKARTSELVQQ- 2144
 Qy 2691 MAEYSPFOLLTFDEPEFMASEIKPQLELOIIOIYIPSTOMQKAFDFDHTTARPPFPV 2740
 Db 2145 -VPLTPVQVRSF-----PEANQREQNH-----NQAFML-----YRENGFAE 2179
 Qy 2741 FYIDPSTSEDAAGLAKACSLVNLHLDIFRTVFAEASGELYQ-----VLSCLDLPQVI 2796
 Db 2180 RIVE-----KVFRKLTEHHDALDMVWYKNGDIIQHNRLGLEDVSFVLYVYDL 2226
 Qy 2797 ETEDNINTATNEFDEFAKPEVRLCHPLRIITIKQTKSMRVIMRISHALYDGLSLEHV 2856
 Db 2227 KTERNLEKTVQIATNTQKDISSEGKMKLCVPTTEGDHLLTAIHHLLVDGVSWRLP 2286
 Qy 2857 RKLHLY-----NGRSLPPHPSRYMQYADRGSGHG-----FWRDVQIONTPTIL 2904
 Db 2287 EDFEAYGOALQGRPIELGYKTSYKTFSEKLAEVANSKLLKEQYVREISKGMFLP 2346
 Qy 2905 SDDTWDGNDATCALHLSKIVNIPSOVLRGSSNII-TQATVFNACALVLSRESKDV 2963
 Db 2347 KHROAHNDYENSRTLSISQTEQTEQLLKEAHKAYNTQINDLLLTALLIASRQLTGENR 2406
 Qy 2964 VFGRIVGRQGLPEYQDI-----VGPCTNAPVRAHIESDYNQLLHDIDQYLLSLPH 3018
 Db 2407 L--KILMGHGRDDILQDVDTITRVGWFTAMPYFIDLEADLSDVMKIVKETLRKIPN 2464
 Qy 3019 ETIGFSDLK 3027
 Db 2465 NGIGYGILK 2473

RESULT 10
 HTSL_COCCA STANDARD; PRT; 5217 AA.
 AC Q01886;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE HC-toxin synthetase (EC 6.3.2.-) (HTS).
 GN HTSL.
 OS Cochliobolus carbonum (Bipolaris zeicola).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
 OC Pleosporales; Pleosporaceae; Cochliobolus.
 OX NCBI_TaxID=5017;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-ATCC 90305 / SB111;
 RX MEDLINE=93100328; PubMed=1281482;
 RA Scott-Craig J.S., Panaccione D.G., Pocard J.-A., Walton J.D.;
 RT "The cyclic peptide synthetase catalyzing HC-toxin production in the
 filamentous fungus Cochliobolus carbonum is encoded by a
 15.7-kilobase open reading frame."
 RL J. Biol. Chem. 267:26044-26049 (1992).
 RN [2]
 RP FUNCTION.
 RC STRAIN-ATCC 90305 / SB111;
 RX MEDLINE=20138231; PubMed=10671527;
 RA Cheng Y.-Q., Walton J.D.;
 RT "A eukaryotic alanine racemase gene involved in cyclic peptide
 biosynthesis."
 RT J. Biol. Chem. 275:4906-4911 (2000).
 CC -!- FUNCTION: Non-ribosomal peptide synthetase, able to activate
 proline and AEO (2-amino-9,10-epoxi-8-oxodecanoic acid), and
 epimerize L-pro. Catalyzes the production of HC-toxin: a cyclic
 tetrapeptide. Activates and thioesterifies L-pro, and epimerizes
 it to D-Pro; also uses D-Ala as a substrate but this is epimerized
 from L-Ala by TOXG.
 CC -!- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES.

CC -!- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF HC-TOXIN.
 CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 FAMILY.
 CC -!- SIMILARITY: CONTAINS 4 ACYL CARRIER DOMAINS.
 CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M98024; AAA33023.1; .
 DR HSSP; P14687; 1AMU.
 DR InterPro; IPR000873; AMP-bind.
 DR InterPro; IPR001242; Condensatn.
 DR InterPro; IPR003880; Ppantne_attach.
 DR Pfam; PF00501; AMP-binding; 4.
 DR Pfam; PF00550; pp-binding; 4.
 DR Pfam; PF00668; Condensation; 5.
 DR PRINTS; PR00154; AMPBINDING.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 4.
 DR PROSITE; PS00455; AMP_BINDING; 3.
 DR PROSITE; PS50075; ACP_DOMAIN; 4.
 KW Ligase; Multifunctional enzyme; Phosphopantetheine; Repeat.
 FT REPEAT 249 842 DOMAIN 1.
 FT REPEAT 1854 2452 DOMAIN 2.
 FT REPEAT 3006 3606 DOMAIN 3.
 FT REPEAT 4158 4738 DOMAIN 4.
 FT DOMAIN 762 840 ACYL CARRIER (ACP) 1.
 FT DOMAIN 2384 2450 ACYL CARRIER (ACP) 2.
 FT DOMAIN 3536 3604 ACYL CARRIER (ACP) 3.
 FT DOMAIN 4667 4736 ACYL CARRIER (ACP) 4.
 FT BINDING 803 803 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 2414 2414 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 3568 3568 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 4700 4700 PHOSPHOPANTETHEINE (BY SIMILARITY).
 SQ SEQUENCE 5217 AA; 574637 MW; 0331D9C5400163A5 CRC64;
 Query Match 10.18; Score 1625.5; DB 1; Length 5217;
 Best Local Similarity 22.38; Pred. No. 3.2e-85;
 Matches 736; Conservative 475; Mismatches 1080; Indels 1007; Gaps 125;
 Qy 46 IEAKPCTPFOLDMIDCNALDKQSA-----ICHAVIDVPTDIDISRFALEKWEIYNQ 97
 Db 1337 LEGYVPGSQMDALFLSQSKSQDCGAYEVDFTRVATSLQNSQPAVDICLVENAKDTVAL 1396
 Qy 98 TPALRAFAFTSD---SGKTSQVILKDSFVFSMMCMWSSSSSPDEV---VRDEAAAA---A 147
 Db 1397 HAALRTVILESSLPATGILHGVLR-----SHDPDIDVLDRVDTAAITLDS 1444
 Qy 148 SGPCRNFEVLLEDQTKKCOLVWF-----SHALVDVTFQQRVLSRVFAAYKHEK 197
 Db 1445 YPPTTEGIALIKRPPHRLICTTIEGSLVKFKVNHILVFDGMDTKTIQDLSKAYTCRH 1504
 Qy 198 DTHRPETPSS--DAT-----DTDSQSVSVSVSMSCEDNAVSATHEFWQ 237
 Db 1505 SNKLDPHSESKLHDGTGNGRTPKPLAEFIRYIRDPQKQDSI-----NYWK 1551
 Qy 238 THLDNLNASVFPPLHSDHL-----MVPNTTAEHRTFPPLSQKALSNAICRT 285
 Db 1552 NALRGATTCSPPPLFDQITSEKAMPROSASVPIPLCVDSKELSKTLANLGTWTSTMFOT 1611
 Qy 286 ALSTLLSRYTHSDEALFCAVTE--OSLPDKHYLAGDYQTVAPLRVHCQSNLRASDVMDA 344
 Db 1612 WVALVRIYSONGSQSVFGLTSGRDAPVDGSDSAVGNF--IA-----MLVC 1655
 Qy 345 ISSYDDRLGLHAPFLGRDIRNTGNGSA--AC----DFQTVLLVTDGSHVNGINGFLQOI 399
 Db 1656 FFDFDDGGVHTVADMARKIHNASNSISHQACSLAEIQDALGLSTSTPLFTNTATYLPK- 1714

QY 400 TESSHEMPCNNRA-----LLHCOMESSGALLVAYDHNV--ID 436
Db 1715 -----RPNVKGPEHHLCCFEELSMSPDTEFLTLFVEPTQESNEVSAHLDFKLSYIS 1768
QY 437 SLOTRLLQOFGHLI-KCLOSPDLJLSSMAEVLNMTEDYDRAEIESWNSQPLEVQDTLIHHE 495
Db 1769 QAYATSIATVAHLSLVHDP--YRALNTLPVSEHDTAIRSNDHFLPPATECIEHET 1826
QY 496 MLKAVSHSPKTAIQAWDGTWYSELNYSRLAVHIKSLGRAOQAIIIPVPEKSKWVI 555
Db 1827 FRKVVVEHPQREACISWDGSLTYAELSDLSQRLSIHLVSLGIKVGTK-IPICEKSMWTI 1885
QY 556 ASMLAVLKSGNAFTLDPNDPPARTAAVVTQTRATVAL-----TSKLHR-ETVOKLVG--- 607
Db 1886 VTILAVQAGGVFVLEPHGPERLSGIIKQVQAEILLCSPTSRMGALQNTSTOMGTET 1945
QY 608 RCVVVDELLQ--VSASODFSLSKSQLAVYFTSGTGDPKGMIEHRAFPSSCALKF 665
Db 1946 KIVELEPEFIRSLPLPKPNHQPMVGLNDDLYVFTSGTGVPKGAVATHQAYATGIYEH 2005
QY 666 GASLGINS---DTRALQFTHFAGACLEIMTTLINGCVCIPSDDDRMNS-IPSFINRY 721
Db 2006 AVACGMTSLGAPPSLQFASYSFASIGDIFTTAVGGCLCIPREEDRNPAGITTFINRY 2065
QY 722 NYNNMMATPSYMGTFSPEDVPGLATLVLYGEQMSSSVNAIWPALKOLLNGYQSESSIC 781
Db 2066 GVTWAGITPSLALHLPDPAVPTLKALCVAGEPLSMVSVVWMSKRLNLINNYGTEATVAC 2125
QY 782 FASNM---STEPNNMGRVAGHSWIDPNDINRLVPIGAVGELVIESPGIARDYIVPPPP 838
Db 2126 IANOVCTTITVSDICRGYRATVWVQPDNHNLSLVPIGAVGELIIEGSLCRGYL-NDPE 2184
QY 839 EKSPFTDTPSWYPANTFPDGAKYRTGDLARYASDGSIVCLGRIDSQVKIRGORVELGA 898
Db 2185 RTAEVIRPSW--LHDLRPNSTLYKTGLVRYSDGKIIFIGRKDTQVMKNGQREFELGE 2242
QY 899 IETHLRQOM-PDDLTVVATKRSQSANSTSLIAFL-IGSSVFGNRPSPDAHILD----- 950
Db 2243 VEHALQLOLDPDGP IIVDLLKRTQSGEPDLTIAFLFVGRANTGNSDEIFTATSTSSL 2302
QY 951 -----HDATKAINIKLEQVLRHSPISFYICMLE--LPRTATGIDRRRLRIM-- 996
Db 2303 SEESTVIKKLODAQRAM-----EVLPLFWVPQAYI-PIEGGLPTAAGKIDRMLKICE 2356
QY 997 --GKOILDQOTGAI---VOQAPAPIPVFADTAA-KLHSIWVQSGLIDPATNVVG--ATF 1048
Db 2357 PFNRNDLISFTSKALSTSVDKDAET-----TDTVEDRLARIWEKVLGVK---GVGRESDF 2407
QY 1049 FELGGSNITAIKWNVNARSVMGLKVSNIYQHPTLAGISAVVG-----DPLSYTL---IP 1101
Db 2408 FSSGNSMAAIALRAEPAQRSGFTLFVADIFTNPRADMAKLFSGHGSVSPSSSTLRKVP 2467
QY 1102 KSTHEG-----PVEQSYSGO-----RWFLDQ 1123
Db 2468 ISSLOKRSGLQTAAPVNSGSPVRCCKENIDCPVAFYEBEGSPDTQLKEASRICGISS 2527
QY 1124 LVGSLWYLIPVAVMRGPNVVDALRALAALE-----QRHETLRTTFEDQDGV----- 1172
Db 2528 RSIEDVFPCTP-----MGEALVALSLIPGAQASVALHAAFELRPLGLDRNFR 2574
QY 1173 -----GVQIVHEKLEEMKVIDLGSDDLDPFVNLQEQOTTPN 1210
Db 2575 SAWESTVKAQPIILRSRIISGSSGVVVTSAATDIPQLDVG--LDTF--LEQLOQVGPA 2630
QY 1211 LSSEAGWRATLLRGEDDHILITVMHHIISDGWSIDVLRRLDNLNOLYSAALKDSDPLSAL 1270
Db 2631 PCAPLERLAFVYSKADDCDYFVLSAHHAIDYDGSNLNLIWSQVLYATNG----- 2679
QY 1271 TPLPTOYSDFAKWQKDFTEQBK--OLNYYKKOL--KDSPPAKIP-----TD 1313
Db 2680 -ELPPPGPSFKHARNLNLVQSKLSDSEDFWRKLLVKPDQESFRFPDVPVGHKPAKTCITN 2738
QY 1314 FARPALLSDAGCVHVTIDGELYQSLRAFACNEHNTTSFVLLAALFAAAHYRLTAVEDAVI 1373

Db 2739 FHEPFSMQSKIG-----TTANTCINAAMAITLAQYSSNKTVMF 2776
QY 1374 GTPIANRRP--ELEDIIGCFVNTQCMRINIDHHDFTGLINOVKATTTAAAFENEDIPFE 1431
Db 2777 GVTLWGRDFPMIDIEHTGPTIVTVPROVNVIPESSVAEFLQDLQKSLAVVLPHOGLGLH 2836
QY 1432 RVVSALQPSGR---DLSSTPLAQLIFAVHSOKDLGRFKQGLSVPPVPSKAYTRFDMFEH 1488
Db 2837 R-IQALGPITARQACDFST-----LLVNHGSSISWSELEAADIVPPLSSDLAYPMV 2889
QY 1489 LFQE---TDSLKGSVNAFDELFKMETVENVVRVFFELLRNGLOSSRTPVSIPL-----L 1539
Db 2890 VEVENASSDTLDIRVHSDPDCEIEVQLERLMEQF---GHNLOTLCLRAASFDEGKRIAE 2945
QY 1540 TDCIVT--LEKLDVLANVKHVDYPRESSLA--DVFOQVSAYPDSLAVVDSSCLTYTFLD 1595
Db 2946 MDDTATHTLETFSWNSRVKSDPVAIAIVHKLLEETAQSQPAESAIVAHGQSLYMQMD 3005
QY 1596 ROSDILAGMLRRSRM--PAETLVAVFAPRSCETIVAFFGVGLKANLAYLPDVRSPARVQ 1653
Db 3006 RCADVLARQIRKTNMISAQSPFVCIHLLRSATAVVSMVLAVLKAGGAMPVDISQPSRLQ 3065
QY 1654 DILSGLSGPTIVLIGHDTAPPD-----TEVTNVFVRIRDALNDSNADGFEVIEH 1703
Db 3066 NLIEE-SGAKLVL---TLPEANALATLSGLTKVIPVLSLSELVQQITDNTTKKEYCKS 3120
QY 1704 DSTKPSATSLAVLYLTSGSTGRPKGVMIHRVIRTVTSGCIPNY---PSETMAHMATI 1760
Db 3121 GQTDPS--SPAYLLYTSGTSGKPKGVWHEHRANSLGFT--CHAEYMGFNCTRLQLFSSL 3176
QY 1761 AFDGASYEISALLFGRTLVCVDMYMTLDARALKDVFREHVNAAHSVTSSSODVPLRVP 1820
Db 3177 MFLSILEIWAIVLYAGCL-----FIPSDKERYNNLQDFTINDINTVELTPTSGKLLNP 3231
QY 1821 RRLSRTLMEFFLVVVTSTAPDALDAQLYGVOCYNGYGTENGVMSTIYPIDSTESFIN 1880
Db 3232 KDLPN--ISFAGPIGEPMTRSLIDAWTL-PGRRLVNSYGPTEACVLVAREISPTAP--H 3286
QY 1881 GVP---IGRALNNSGAYVVDPEQ-QLVGVIGMGMELVVVTGDLARGY--SDKALDENRVHI 1935
Db 3287 DKPSSNIGHAL-CANIWVVEPORTALVPVIGAVGELCIEAPSLARCILANPERTEYSEPT 3345
QY 1936 TV-NDQTVK---AYRTGDRVYRIGDGLIEFFGFRMDTFQKIRGNRIESAIEAALLRDS 1991
Db 3346 VLDNMWTKKTRVYRTGDLVRY-ASDGTLDLGRKDGQIKLRGORTLGEIE----- 3396
QY 1992 VRDAAVVLQONEDQAEIILGFVVADHDHSENDKGSANQVGEHQDHFECSNYSDIGEIDP 2051
Db 3397 -----HHIRLMSD-----DP 3407
QY 2052 STIGSDFKGTSMYDGSQIDFDBMHEWLGETTTLHDNRSLGNVLEIGTCGMLFLNLS 2111
Db 3408 -----RFHE----- 3411
QY 2112 RLSEYVGLPEFSAAAFVNKATESIPSLAGAKVQV-GTATDQGVDDLHPDLVLVNSVI 2170
Db 3412 -----ASVQLYNPATD-----PD----- 3424
QY 2171 QYFPPSEYLAETADTLIHLPNVORIFFGVRSQATNEHFLAARAIHTGKNATKDDVRQK 2230
Db 3425 -----RDATVD----- 3430
QY 2231 MAELEDMEBELLVEPAFFTSLKDRFPGLVEHVEILPKNMBEAVNLSAYRAYAAVHVGRSL 2290
Db 3431 ----- 3430
QY 2291 GDELVLPEKDWIDFQANQLNOKSLGDLKSSDAAIMAVSKIPFEITAFEROVVASLNS 2350
Db 3431 -----VQMPREPLYLAGL----- 3442
QY 2351 NIDEWOLSTIRSAEGDSSLSVPDIFRIAGEAGFRVEVSAROWSONGALDAVFHCCSQ 2410


```

Db 3443 -----VLDL-----VSLMRSDS-----MHVNI 3461
QY 2411 GRTLVNFDHLLRGSDLLTNRLQRLQNRRIAEVRLRLSLPSYMPISNIVVLKMP 2470
Db 3462 ANISENLQT-----LVTELKSLRGVLPHYWVPLHFVAVSRUP 3499
QY 2471 LNANGKVDKRLSRRAKVPKQTAAPLP-TFFPSEVEVLCEATEVFGMKVDIITDHF 2529
Db 3500 TGSSGKLDAFV--RACL--RELTAFLDGNFP--KVEQVLTN-----3536
QY 2530 NLGGHSLATKLISRIDQRKVRITVKDVFEDHVFADLASVIRGLG--LQOPVSDGOG 2586
Db 3537 -----ESVLKQWGTVLAMDPHSITQRG 3558
QY 2587 QDRSAHMAPRTEFAILCDEFAKVLGFQVGIITDNFELGGHSLMARK---LAVRIGHRLD 2643
Db 3559 -----DDFESLGGSSISAMRLVGLARSSGHKLQ 3586
QY 2644 TTVSVKDVDPHVPVLFOLATALONLVOSKNEIVGGREMAEYSP-----FOLLTFEDPE 2696
Db 3587 H-----EDIFMCPRLADMAGQI-SFQOE-----ASVSPTTSPTIKFDLLDCEVD 3630
QY 2697 EFMAEIKPOLEL-QEIIODIYPSTOMQAKFLDHTTARPRPVPFYIDFPSTSEPDAG 2755
Db 3631 E-VIDHLPOLDMKNKELIEDVPCTPLQESLM--AATARHGEAYTM-IOSITVLAALQ 3686
QY 2756 LIKACSLNVHLDIFRTVFAEASGELYQVVLSCLDLPIQVIEFDNIN----TATNEFLD 2811
Db 3687 LKAMDWFRDFEVLRT-----RIALGPSQALQVVKHEELSWESFPIQSFKD 3736
QY 2812 EFAKEPVRGLHPLIRITIKQ-----TK-----SMVIMRISHALVDGLS 2851
Db 3737 HEYRS-LGCGKPLARLAVITQALDKOPTSHGTREARTKNSQDTVMVVGCAHHSIYDAH 3795
QY 2852 LEHVVRKLHMLYNG-----RSLPPHOFYSRYMOYTABDRESGHC--FWROV 2895
Db 3796 LSMWRRLYREFIGSQADGILEAETSRSRGVVP---FKSYVEKLLRGKDNESLLEWKEK 3852
QY 2896 ----IQNTPMTISDDTVVDCGNATCKALHLSKIVNIPSOVLGRSSNIITQATVFNACAL 2952
Db 3853 LRGVSSSQPPASWPRVLEHOPSATQTL-ITK-VSLPTSSRKLG--ATVATVAYAAWAL 3908
QY 2953 VLSRESDSKDVFGRTVGRGRO---GLPVEYQDIVGCTNAPVRAHIESSDYNQLLHD 3007
Db 3909 TIAHYTADPDVFGATLSGRETMAGSISHPESIAGTPIITVPLRII---DFQTVVSD 3963

RESULT 11
SRFL_BACSU
ID SRFL_BACSU STANDARD; PRT; 3587 AA.
AC P2706;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Surfactin synthetase subunit 1.
GN SF6A OR SRFA OR SRFA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=93181186; PubMed=8441623;
RA Fuma S., Fujishima Y., Corbell N., D'Souza C., Nakano M.M.,
RA Zuber P., Yamane K.;
RT "Nucleotide sequence of 5' portion of srfA that contains the region
RL required for competence establishment in Bacillus subtilis.";
RL Nucleic Acids Res. 21:93-97(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=93360813; PubMed=83355609;
RA Cosmina P., Rodriguez F., de Ferri F., Grandi G., Perego M.,

```

```

RA Venema G., van Sinderen D.;
RT "Sequence and analysis of the genetic locus responsible for surfactin
RT synthesis in Bacillus subtilis.";
RL Mol. Microbiol. 8:821-831(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124189; PubMed=8969502;
RA Yamane K., Kumano M., Kurita K.;
RT "The 25 degrees-36 degrees region of the Bacillus subtilis
RT chromosome: determination of the sequence of a 146 kb segment and
RT identification of 113 genes.";
RL Microbiology 142:3047-3056(1996).
RN [4]
RP SEQUENCE OF 1-460 FROM N.A.
RX MEDLINE=91154134; PubMed=1847909;
RA Nakano M.M., Magnuson R., Myers A.M., Curry J., Grossman A.D.,
RA Zuber P.;
RT "srfA is an operon required for surfactin production, competence
RT development, and efficient sporulation in Bacillus subtilis.";
RL J. Bacteriol. 173:1770-1778(1991).
RN [5]
RP SEQUENCE OF 1-38 FROM N.A.
RX MEDLINE=91358326; PubMed=1715856;
RA Nakano M.M., Xia L., Zuber P.;
RT "Transcription initiation region of the srfA operon, which is
RT controlled by the comp-comA signal transduction system in Bacillus
RT subtilis.";
RL J. Bacteriol. 173:5487-5493(1991).
RN [6]
RP SEQUENCE OF 1-64 FROM N.A.
RC STRAIN=168;
RX MEDLINE=95219080; PubMed=7704255;
RA Fujishima Y., Yamane K.;
RT "A 10 kb nucleotide sequence at the 5' flanking region (32 degrees)
RL of srfA of the Bacillus subtilis chromosome.";
RL Microbiology 141:277-279(1995).
CC -!- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME ABLE TO
CC ACTIVATE AND POLYMERIZE THE AMINO ACIDS LEU, GLU, ASP AND VAL.
CC -!- ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
CC -!- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -!- PATHWAY: Cyclic peptide antibiotic surfactin biosynthesis.
CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D13262; BAA02522.1; -
CC EMBL; X70356; CAA49816.1; -
CC EMBL; D50453; BAA08982.1; -
CC EMBL; M59939; BAA22815.1; -
CC EMBL; M64702; BAA22816.1; -
CC EMBL; D30762; BAA21034.1; -
CC EMBL; Z99105; CAB12142.1; -
CC FIR; S35517; S35517.
CC HSSP; P14687; IAMU.
CC Subtilisin; BG10168; srfAA.
CC InterPro; IPR000873; AMP-bind.
CC InterPro; IPR001242; Condensatn.
CC InterPro; IPR003880; Ppantne_attach.
CC Pfam; PF00501; AMP-binding; 3.
CC Pfam; PF00550; pp-binding; 3.
CC Pfam; PF00668; Condensation; 4.
CC PRINTS; PR00154; AMPBINDING.
CC PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
CC PROSITE; PS00455; AMP_BINDING; 3.

```

DR PROSITE: PS50075; ACP_DOMAIN; 3.
KW Ligase; Antibiotic biosynthesis; Complete proteome.
KW Multifunctional enzyme; Repeat; DOMAIN 1 (GLU-ACTIVATING).
FT REPEAT ? 1047
FT REPEAT ? 2086
FT REPEAT ? 3114
FT REPEAT ? 3114
FT DOMAIN 976
FT DOMAIN 2015
FT DOMAIN 3043
FT BINDING 1006
FT BINDING 2045
FT BINDING 3073
FT BINDING 3073
FT CONFLICT 146
FT CONFLICT 151
FT CONFLICT 165
FT CONFLICT 281
FT CONFLICT 281
FT CONFLICT 460
FT CONFLICT 540
FT CONFLICT 540
FT CONFLICT 562
FT CONFLICT 639
FT CONFLICT 644
FT CONFLICT 644
FT CONFLICT 647
FT CONFLICT 649
FT CONFLICT 1026
FT CONFLICT 1065
FT CONFLICT 1132
FT CONFLICT 1133
FT CONFLICT 1164
FT CONFLICT 1164
FT CONFLICT 1458
FT CONFLICT 1458
FT CONFLICT 1850
FT CONFLICT 1850
FT CONFLICT 1894
FT CONFLICT 1894
FT CONFLICT 1911
FT CONFLICT 1911
FT CONFLICT 1973
FT CONFLICT 1973
FT CONFLICT 2052
FT CONFLICT 2052
FT CONFLICT 2218
FT CONFLICT 2218
FT CONFLICT 2265
FT CONFLICT 2265
FT CONFLICT 2291
FT CONFLICT 2291
FT CONFLICT 2349
FT CONFLICT 2349
FT CONFLICT 2428
FT CONFLICT 2428
FT CONFLICT 2611
FT CONFLICT 2611
FT CONFLICT 2685
FT CONFLICT 2685
FT CONFLICT 2735
FT CONFLICT 2735
FT CONFLICT 2896
FT CONFLICT 2896
FT CONFLICT 3025
FT CONFLICT 3025
FT CONFLICT 3096
FT CONFLICT 3096
FT CONFLICT 3271
FT CONFLICT 3271
FT CONFLICT 3316
FT CONFLICT 3316
FT CONFLICT 3451
FT CONFLICT 3451
FT CONFLICT 3483
FT CONFLICT 3483
FT CONFLICT 3487
FT CONFLICT 3487
FT CONFLICT 3493
FT CONFLICT 3493
FT CONFLICT 3495
FT CONFLICT 3495
FT CONFLICT 3499
FT CONFLICT 3499
FT CONFLICT 3507
FT CONFLICT 3507
FT CONFLICT 3587
FT CONFLICT 3587
FT SEQUENCE 3587 AA; 402424 MW; ALE2DABF93EDE3A CRC64;
Query Match 10.0%; Score 1613; DB 1; Length 3587;
Best Local Similarity 21.6%; Pred. No. 8.9e-85;
Matches 696; Conservative 483; Mismatches 1101; Indels 946; Gaps 108;
QY 82 IDIRFALAWKEIVNOTPALRAFAFTSDSGKTSQVI-----LKDS-----FVF 124
DB 42 IDVLVEQAIQEFIRNDAMRLRLDENGEPVQVISEYRVPVDIKHTDITDTPNATEFIS 101
QY 125 SHMCHSSSSSPDEVVDEAAASGPRCN--RFVLLDMOTKKCOLVWTFSS---HALVDV 179
DB 102 QMS-----REETKPLPLDYCDLDFRSLF---TIKENVWFYANVHVHISDG 145
QY 180 TFQQRVLSRVFAAYKHEKDTHRPETSSDARDTDSQSVSVSMSCEDNAVSAATH-----234
DB 146 ISWNILGNAMHIYL-----ELASASETEGISHSHFIDHVLSEQYQAQSRFEKDKA 197

QY 235 FWOTHNLNDL-----NASVFPHL--SDHLMVNPNTTAAEHRIITFPLSQKALSNSAICRT 285
DB 198 FNNKQFESVPELVSLKRNASAGSLDAERFSKDVPEALHQIILSFCEANK--VSVLSVFQS 256
QY 286 ALSILLSRTHSDEALFGA--VTEQSLPDKHYLADGTOTVAPLVRVHCOSNLRASD--VMD 343
DB 257 LLAAYLYRVSGONDVVVTGTFMGNRTNAKEKOML--GMFVSTVPLRNTIDGGOAFSEFVKD 314
QY 344 ATSSYDDRLGHILA--PGL--RDIRNTGDN-----GSAACDFOT-----VLLVTDGSHV 388
DB 315 RKMDLMKTLRHOKYPYNLLINDRETSSLTKLFTVSLQVYVQWQKEEDLAFLEPIFS 374
QY 389 NNGINGFLOQITESHFPCNNRALLHQMMESSGALLVAYYDHNVIDSLQTLRLQFQG 448
DB 375 GSGLND-----VSIH-----VKRDWDTGKLTIDED 399
QY 449 HLKICLOSPDLSSMAEVNLMTEYDRAETESNNSQPLEVOD--TLI---HHEML-----497
DB 400 Y-----RTDLFSREEINNICERMITMLENALTHTPEHTIDELTISDAEKEKLLARAG 452
QY 498 KAVSHSPTKT-----AIQAWDGDWTYSELDNNVSSSLAVHIIKSLGURAOQ 541
DB 453 KVSYSYKDMTIPELFOEKABLLSDHPAVVFEDRTLSYRLHQSARIAVNLKOKGV--GPD 511
QY 542 AIIPVYFEKSWKVIASMLAVLKSNGAFTLIDPNDPARTAQVVTQTRATVALT--SKLHR--599
DB 512 SPVAVLIERSERMITAIMILKAGGAYVPIDPGFAERIQYILEDCGADFILTESVAAP 571
QY 600 ETQVKLVGRVVDDELLOSASDDFSSLTQSODLAYVIFTSQDGPDKGIMIEHRAFS 559
DB 572 EADAELIDIDQAIIEGAESLNAD-----VNARNLAYIYTSQGTGRPKGVMIHROVH 625
QY 660 SCALFKGASLGINSDFRALQFGTHA---FGACLEITMTTLINGCVCIFSDDDDMN--SI 714
DB 626 HLVESLQOII--YQSGSQTLMALLAPFHDASVKQIFASLLIGQTLIYVPKKTVTNGAAL 684
QY 715 PSFINRYNNMMHATPSYMGTFSP--DVPLAT--LVLYGEOQSSSVNAIWAFL-----766
DB 685 TAYYRKNSTEATDGTFAHQLMLAAAGDFGLKHLMLIGGEGLSVV---ADKLLKLFK 740
QY 767 -----QLLINGYQOESSSICF--ASNSTEPNN-----MCRAYGAHS--WVIDPNDIN 810
DB 741 EAGTAPRLTNVYGPTE---CVDASVHPVIPAENAVQSAVYPIGKALGNRLYLDDQK--G 795
QY 811 RLVPIGAVGELVIESPGIARDYIVPPPEKSPFETDIPSWYPANTEPDGAKLYRTGDAR 870
DB 796 RLQPEGVAGELYIAGDVGVRGYLHLPDELTEEFQD-----PVPDGRMYRTGDVVR 847
QY 871 YASDGSIVCLGRIDSQVKIRGQVELGAIETHLRQOMPDDLTIVVEATKRSOSANSTSLI 930
DB 848 WLPDGTIEYLGREDQVKVRYRIELGEIETAVI--QAPDPAKAVVLA--RPDQGNLEVC 904
QY 931 AFLIGSSYFCNRPSSDAHILDHDTAKAINKLEQVLRHPSIPSEYICMLLPRTATKIDR 990
DB 905 AYVWQKP--GSEFAPAGLREHAARQ-----LPDYMVPVAFTEVTEIPLTPSKVDR 953
QY 991 RRLRMKGDKILDKQTOGAIVQOAPAPIPVFADTAALKHSIWQSLGIDPATVNVGATFFE 1050
DB 954 RKL-----FALEKAVSGTAYTAPR-----NETEKAIAALWQDVLNVEKA--GIPDNFFE 1001
QY 1051 LGGNSITAKMV--NMARSYGMDBLKVSNIYOHPTLAGISAVV--KGDPLSYTLIPKSPHEGP 1108
DB 1002 TGGHSLKAMTLLTKHKETGIEPLQFLFEHTITALEADHRESKAFATIEPAEKQEH 1061
QY 1109 VEGSYVSGRLWFLDQDLVGLWYLIPIYAVMRGPNVDALRALALEORHETLRTTFED 1168
DB 1062 YPLSLAQOQRTYIVSQFEDAGVGNMPPAAALIEPLDIQKLERAFQGLIRHESLRTSFLV 1121
QY 1169 QDGVGVQIVHEKLEEMKVIDLCG--SDLPDFVFNQEQTPFNLSSEAGWRATLLRIGED 1227
DB 1122 ENSTPRQKTHSDVDNFIEMIERRGSD-----EAIMASFVRTFDLAKRPLFRIGLLGLEEN 1177

```
QY 1228 DHILITVMHHIISDGSIDVLRDLNOLYSAALKDSKDPLSALTPLPIQYSDFAKWQKQ 1287
Db 1178 RHMLFDMHHLISDGSIGIMLEELARIYK-----EQPLDLRLQYKDAVWMSRQ 1228
QY 1288 FIE-OEKOLNWKOLKSDSPA-KIPTDFARPALLSGDACVHVITIDELYQSIRAFPCNE 1345
Db 1229 AAGYKKQAYKWEFAGELPVQLLSYDPRPPVQSEFEGDRVSIKLDAGVKDRNLRLAEQ 1288
QY 1346 HNTTSFVLLAFAAHRYLTAVEDAVIGTPIANRNRPELEDIIGCFVNTQCRINIDHH 1405
Db 1289 NGATLYMVLSAYYTLSSKYTGQDDIIVGTPSAGRNHSDTEGIGMFVNTLAIRSEVKQN 1348
QY 1406 DTGCTLINQYKATTAAFENEDIPFERVVSALQPSRDLSTPLAQILFAVHSQKD----- 1461
Db 1349 ETQTQLSRVRKRVLDAFSDHODYPFEWLVEDLNI-PRDVSRRHPLFDTMFSLOQATEGIPA 1407
QY 1462 LGREFKQGLSESVPVPSKAYVRPOMEHFLFOETDSLKGSVNFADFLFKMETVENVVRVFE 1521
Db 1408 VGDLSL-----SVOETNEFKIAKFDLTVOARETDEGIEDVDYSYKLFKQSTADRLVTHFAR 1463
QY 1522 ILRNGLOSSRTPVSIPLDGTGIVTLEKLDVNLVKNHVDYPPRESSLADVFOTQVAYSPLSA 1581
Db 1464 LLEDAADPEKPISEYKLLSEEAASQIQOFNPGRTYPKDKTIVQLFEEQAANTPDHTA 1523
QY 1582 VVSSCRLTYELDRQSDIILAGWLRRRSMPAETLVAVFAPRSCETIVAFEGVLKANLAYL 1641
Db 1524 LOYEGESLTVRELNRANRLARILSLGAGEGRTAAVLCERSMDMIVSILAVLKSGSAYV 1583
QY 1642 PLDVRSPSARVQDILSGSLGPTIVLIGHDTAPPDIEVTNVEFVRIRDALNDSNADGFEVI 1701
Db 1584 PIDPEHPQRMQHFFRD-SGAKVLLTOR-----KLKALAEAEFEKGVIVL 1627
QY 1702 --EHDSTKPSATSLAYVL-----YTSGSTGRPKGMVMEHRVIRTVTSCGIPNYS 1750
Db 1628 ADEEESHADARNLALPLDSAAANLTYTSGTGTGPKGNIVTHANILRTVKE---TNLS 1684
QY 1751 ETR--MAHMATIAFGASVEIYSALLFGRTLVCVDMYMTLDARALKDVFREHVNAAASH 1807
Db 1685 ITSQDTILGSLNVVDFAMFDMGSLNAGKLVLPKETVLDMARLSRVTERENISILMI 1744
QY 1808 VTSSQDVLRVPRRLS--RTLMEFFLVVTDSTAPDALDAQGLQYQVQYNGYGTENG 1865
Db 1745 TTALFHLVLDNLPACLTSLRKIMFGGERASVEHVRKALQTVG--KG-KLLHMYGPSESTV 1801
QY 1866 MSTIYPIDSTESINGPIGRALNNGAYVVDPEQOLVGTVGMGELVVTGDLGARGYSDK 1925
Db 1802 FATYHPVDELEHTLSYPIGKPVSNTEVYILDRTHVQVAPAGIAGELCVSGEGLVKGYNR 1861
QY 1926 -ALDENRFV-HITVNDQTVKAYTGDVRVYRIGDGLIEFFGRMDTOPKIRGNRIESAEIE 1983
Db 1862 PELTEKFPVPHPTSGE--RMVKTGLARW-LPNGDIEFTRIDHVKIRGQRIELGEIE 1918
QY 1984 AALLRDSVRAAAVLQONEDQAPEILGFVYVADHDHSENDKGOSANQVQGWQDHFESGM 2043
Db 1919 HQL-----QTHDRVOESV----- 1931
QY 2044 SDIGEIDPSTIGSDFKGTWSDYQSGQIDFDEMHEWNLGETRTLHDNRSLGNVLEIGTSG 2103
Db 1932 -----VLAVDQAG 1940
QY 2104 MILFNLDLSRLESYVGLSPRSAAAFVNKATESIPSLAGKAKVQGTATDIGOVDDLHPDL 2163
Db 1941 -----DKLLCAY----- 1947
QY 2164 VILNSVIQPPSPSEYLAETADTLIHLNPNVQIRFFGDRVRSQATNEHFLAARAIHTLGNAT 2223
Db 1948 -----YVGE-----GDISSQEMREH-----AA 1964
QY 2224 KDVROKMAELEDMEELLVEPAFFTSLKDRFGLVEHVEILPKNEAVNELSAYRYAAV 2283
Db 1965 KD----- 1966
QY 2284 VHRVSGLDLVLVPEKDDWIDFQANQNLNOKSLGDLKSSDAAIMAVSKIPFEITAFERQ 2343
```

RESULT 12

```
PPSL_BACSU
ID PPSL_BACSU STANDARD; PRG: 2561 AA.
AC P39845;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Peptide synthetase 1.
GN PPSA OR PPSI.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
```


QY 1054 NSITAKWVN-MARSGVMDLKVSNIOHPTLAGISAVVK-GDPLSYTLIPKSHHEGPVEQ 1111
 Db 995 HSKATALVSKIAKEFDVQPLKDVFAHPTVIREGTSPYBAIKPAEKQETYPV 1054
 QY 1112 SYSOGRLEWFDLDVGLWLVIPYAVRMKGFVNVDALRRALAALEQRHETLRTTFF-DQD 1170
 Db 1055 SSAQKRIYVLOLEDGGTGYNMPAVIELEKLNPERMERAFKELIKHESLRTSFEQDAG 1114
 QY 1171 GVGQVIVHEKSEMKVIDLCGSDLDPPFEVLNQEQT-----PFLNLSSE 1214
 Db 1115 GDPVQRIHDEV-----PFTL-----QTTVLGERTEQEAFAAFKPFDLSSQA 1155
 QY 1215 AGHRATLLRGEDDHILTIYVHHIISDGSIDVLRDLNOLYSAAKDSKDPISALTPLP 1274
 Db 1156 PLFRAQIVTKISDERHLLVDMHHIISDGSVNNILIREFGELYN-----RNLPALR 1206
 QY 1275 IQSDFAKWK-----DOFTEQEKQLNWKOLKDDSPA-KIPTDFARPALLSGDAGCVH 1328
 Db 1207 IQKYDAVWREGTKDAYKTQEA---YWLKLEGELPVLDPADHARPVRPSFAGDKVS 1263
 QY 1329 VTIDGELYQSLRAFCEHNTTSFVVLIAFRAAHYRLTAVEDAVIGTPPIANRPELEDI 1388
 Db 1264 FTLDQEVASGLHLKARENGSTLYMVLAAVTAFLSRLSGQEDIIIVGSPAGRPKHOLEPI 1323
 QY 1389 ICGFVNTQCMRINIDHIDTGTTLINQVATTAFAFENEDIPFRRVVSALOPGRDLSSTP 1448
 Db 1324 LGMFVNTLALTRPEGKPKVQVLOEVRETALEAFHQDPYFEEVLVDKLEL-TRDMSRNP 1382
 QY 1449 LAOLIFAVHS--OKDLGRPKFOGLESVPVPSKAYTR---FDMFEHLFOETDLSLKGVSNF 1502
 Db 1383 VEDAMFILONVEKQDI-----DLREIKVRPANFAHISLFDITLATEISGSCIDMEF 1436
 QY 1503 ADELKMETVENVVRVFEILRNGLOSSRTPVILPLTDGIVTLEKLDVL---NVKHHVY 1559
 Db 1437 STEVFLKATIERWADHFIELHEALS---TPETSLAQINILSDKEKQKIVFEFNKQTVF 1493
 QY 1560 -PRESSLADVPQVQVSAVPSLAVDSSCLTYTDLDRQSDIILAGLWRRMSMAEFLVAV 1618
 Db 1494 AQDIPFHFIFEAKAENPEHIAVDNETEISYRLNERANRLARTLONRKGPKPT-VAV 1552
 QY 1619 FAPRSCTEIVAFRGVLKANLAYLPDVRSPSARVQDILSLSGPTIVLIGHD-----TAP 1673
 Db 1553 LAKRSIDAIVGLAVMKAGGVIPIDAHYPKARIEYLIRD-SCADILLLORELKHLISNS 1611
 QY 1674 PTEVTNVFVRTDALNDSNADGFEVIEHDSTKPSATSLAYVLYTSGSTRGPKGVNIEH 1733
 Db 1612 PESEMSHI-FLDDEGSFEESNC-----NLNLSAPAEPEPVYIYTSGITGAPKGVITY 1663
 QY 1734 RVIIRTVTSCTINYP-----SETRMAHMATIAFDGASYEIIYSALLFGRTL 1779
 Db 1664 Q-----NFTHAALWKOIYELDRKPVRLQIASFPDVFSGDLARTLTNGGTL 1711
 QY 1780 VCDYMTTLARALKDVFREHYNAAASHVTSQQDVPRLVRPRLSRT---LMEFFFLVVD 1836
 Db 1712 IVCPEDETRLEPAEYIKISORT-----VMESTPALLIPWMEVYVRNQFKLPDLIDLILG 1767
 QY 1837 STAPDLDADGLY-----QGVQCNGYNGPTENGVMSTIYPIDSTESFING-----VPIGR 1886
 Db 1768 SDMKVKAQDEKTLTDFRGQSMRIINSYGVTEATIDSSFY-----ETSMGGECTGDNVPIGS 1822
 QY 1887 ALNNSGAYVDPQQLVGVGMGELVVTGDGLARGYSK-----ALDENRFVHIIVNDQ 1940
 Db 1823 PLPNVHMYVLSQFDQIQPIVAGELCIGGAGVAKGYHHPDLTOMKFTENPFY-----S 1876
 QY 1941 TVKAYRTGRVRYRIGDGLIEFFGRMDTQFKIRGNRIEAEIEAALLRDSVSDAAVVLQ 2000
 Db 1877 GERLYRTGDRACW-LPNGTIRLGRMDYQVKINGRYETEEIESVLLQTLCLVREAAVAVQ 1935
 QY 2001 QNEDQAPELIGFVVAHDHSEN-----DKGQSANQVEGQDHFGESMGYSIDGEIDPSTI 2054
 Db 1936 HDKNGQAGLAAYTVPS-DVNTNALRAALTKELPAYMIPAYLIPLVNPLMTLNGKLDORNAL 1994
 QY 2055 GSDFKGWTSMYDGSQIDFDEHMEWLGETTTRT-----LHDNRSLGNVLEICTGSGMILFNL 2109

Db 1995 PAPNNVLSRPTAPVNDLQKTMAYIWEDVLSMSRVGIHD-----SFFELG-GDSIKALQV 2048
 QY 2110 DSRLESVYVLEPSRSAAF--VNKATESIPSLA-----GRAKYQVGTATDIGQVDD 2158
 Db 2049 AARLAAGWSMTIRDLFRYSTIQELGHIPTLASQADGQPAEGAEALTPIQRRFFGQVHA 2108
 QY 2159 LHPDLVVLNSVIOYFSPSEYLAETADTLI-----HLPNVORIFFCDVRSQAT----- 2205
 Db 2109 FH---YHYNQSVMLFSEKGFNANALHLAKRITHEHDAIRMFORDQNGHVIOFNRGINH 2165
 QY 2206 NEHPL-----AARAIHTLGKNATKDDVQKMAELEDMEELLVEPAFTTSKDR 2254
 Db 2166 KDHELFGLYISDWTWKASLERAHLDKLAABETVIOQSKMNV--KGPLLQAGLFKTAEGDH 2223
 QY 2255 FPGVHEVHEILLPKNMEAVNELSVRYAAVHVVRGSLGDELVLPEKDDWIDFOANOLNQ- 2313
 Db 2224 LLTALHHLVLDGVSWRILLEDLAAAYQOALEK-----EQLPPKTDLSYLSY-ADGLTQI 2277
 QY 2314 -----KSLGDLK-SSDAAIMAVSKIPFEITAFERQ- 2343
 Db 2278 AESKQLLSEKTYWQITLDAHTAFLPKDIENVDPKLOMNSDAAAFVLSGDWTEKLLFTQO 2337
 QY 2344 -----VVASLNSNIDEWQ-----LSTIRSSAEGSSLSVP--DIFRIAG 2380
 Db 2338 AYGTDANELLTALGMALSEWTHQDQIVISTEGHREG-----HVPNIDISRTVG 2387.
 RESULT 13
 ACVS_CEPAC STANDARD; PRT; 3712 AA.
 ID ACVS_CEPAC
 AC P25464;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22; Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Delta (L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase
 DE (SC 6,---) (ACVS synthetase) (ACVS).
 GN PCBAB.
 OS Cephalosporium acremonium (Acremonium chrysogenum).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; Hypocreaceae; mitosporic Hypocreaceae; Acremonium.
 OX NCBI_TaxID=5044;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=91177827; PubMed=1706706;
 RX Gutierrez S., Diez B., Montenegro E., Martin J.F.;
 RT "Characterization of the Cephalosporium acremonium pcbAB gene
 encoding alpha-aminoadipyl-cysteinyl-valine synthetase, a large
 multidomain peptide synthetase: linkage to the pcbC gene as a cluster
 of early cephalosporin biosynthetic genes and evidence of multiple
 functional domains.";
 RT J. Bacteriol. 173:2354-2365(1991).
 RN [2]
 RP PARTIAL SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=ATCC 11350;
 RX MEDLINE=91168300; PubMed=2076552;
 RA Hoskins J.A., O'Callaghan N., Queener S.W., Cantwell C.A., Wood J.S.,
 RA Chen V.J., Skatrud P.L.;
 RT "Gene disruption of the pcbAB gene encoding ACV synthetase in
 Cephalosporium acremonium.";
 RL Curr. Genet. 18:523-530(1990).
 CC -!- FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF THE TRIPEPTIDE
 CC ACV ARE ACTIVATED AS AMINOACYL-ADENYLATES WITH PEPTIDE BONDS
 CC FORMED THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER
 CC INTERMEDIATES.
 CC -!- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTHETHEINES
 CC (POTENTIAL).
 CC -!- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF PENICILLIN AND
 CC CEPHALOSPORIN.
 CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 CC FAMILY.
 CC -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
 DR PIR; A38531; YCEVC.

Db 2417 GVLNPEP---TNLPSSRDSGNSLHGLFEDIVASTPDRIATACTRSLYSSELNERANQL 2473

Qy 1601 LAGWLRRSRMPAETLVAVFAPRSCETIVAFGVLKANLAYLPLDVRSPSARVQDILSGLS 1660

Db 2474 VHLIISASIVADDRIALLDKSIDMVIALLAVKAGAAVPLDPTYPSQRTLEILESS 2533

Qy 1661 GPTIVLGHDTAPPDIEVTNVEFVRI--RQALNDSNADGEVIEHDSKTSATSLAVLY 1718

Db 2534 ARTLITTRKHT-PRGGTVANVPSVYLDSPETLACLNQSKENPPTSTQKPS--DLAYVIF 2590

Qy 1719 TSGTGRPKGVMEHVIITVTSGCIPNYPSETRMAH---MATIADFQASVEIYSALL 1774

Db 2591 TSGTGRPKGVMEHVIITVTSGCIPNYPSETRMAH---MATIADFQASVEIYSALL 1774

Qy 1775 FGRTLCVDYMTILDARALKDVFREHVNAASHVTSQQDVP-L-RVPRRLSRTLMFFFLV 1833

Db 2650 GGNKLI-IPPEGLTHEAFYDGRREKLSYLSGTPSVLQIELSRLPH-----LHMVTA 2703

Qy 1834 VTDSAPDALDAQLOQVOCYNGYGTENGVMSTIYPIDTSFNGVPIGRAL-----N 1889

Db 2704 GEEFHASQFEMKRSQFAG-QINNAYGITE---TTVYNIITT--FKGDAPFTKALCHGIP 2756

Qy 1890 NSGAYVDPPEQQLVGIGVMEGLVVTGDLARGY-SDKALDENRFV-----HITVNDOT 1941

Db 2757 GSHVYVNDRLQRPVFNANGVELYLGDCIARGYLNQDALTNERFIPNPFYEPKQASDRP 2816

Qy 1942 VKAYRTGDRVRYRIGDGLIEFFGRMDTQFKIRGNRIESAIEAALRDSVRAAVVLOQ 2001

Db 2817 QRLYKTKDGLVRFER-GPHILEYLGKQKQVLRGFRLELSEVRDAVLAISAVKEAAVPIKY 2875

Qy 2002 NED 2004

Db 2876 DED 2878

RESULT 14

ACVS_NOCILA STANDARD; PRT; 3649 AA.

AC P27M33

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Delta (L-alpha-aminocadipyl)-L-cysteiny-D-valine synthetase

DE (EC 6.---) (ACV synthetase) (ACVS).

GN PCBA.

OS Nocardia lactandurans.

OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;

OC Actinomycetales; Pseudonocardineae; Pseudonocardaceae;

OC Amycolatopsis.

OX NCBI_TaxID=1913;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=VAR LC 411;

RX MEDLINE=92065808; PubMed=1956290;

RA Coque J.J.R., Martin J.F., Calzada J.G., Liras P.;

RT "The cephamycin biosynthetic genes pcbaA, encoding a large

RT multidomain peptide synthetase, and pcbaC of Nocardia lactandurans are

RT clustered together in an organization different from the same genes

RT in *Acromonium chrysogenum* and *Penicillium chrysogenum*."

RL Mol. Microbiol. 5:1125-1133(1991).

CC -1- FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF THE TRIPEPTIDE

CC ACV ARE ACTIVATED AS AMINOACYL-ADENYLATES WITH PEPTIDE BONDS

CC FORMED THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER

CC INTERMEDIATES.

CC -1- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.

CC -1- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF PENICILLIN AND

CC CEPHALOSPORIN.

CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME

CC FAMILY.

CC -1- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed, usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X57310; CAA0561.1; -

DR PIR; S18268; S18268.

DR HSPP; P14687; IAMU.

DR InterPro; IPR000873; AMP-bind.

DR InterPro; IPR001242; Condensatn.

DR InterPro; IPR003880; Ppantne_attach.

DR InterPro; IPR000379; Ser_estrs_site.

DR InterPro; IPR001031; Thioesterase.

DR Pfam; PF005501; AMP-binding; 3.

DR Pfam; PF00658; Condensation; 3.

DR Pfam; PF00975; Thioesterase; 1.

DR PRINTS; PR00154; AMPBINDING.

DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.

DR PROSITE; PS00455; AMP_BINDING; 1.

DR PROSITE; PS00075; ACP_DOMAIN; 3.

DR Ligase; Antibiotic biosynthesis; Multifunctional enzyme;

Repeat; Phosphopantetheine.

KW REPEAT 401 861 DOMAIN 1 (ADIPATE-ACTIVATING);

FT REPEAT 1014 1937 DOMAIN 2 (CYSTEINE-ACTIVATING).

FT REPEAT 2079 2985 DOMAIN 3 (VALINE-ACTIVATING).

FT DOMAIN 788 857 ACYL CARRIER (ACP) 1.

FT DOMAIN 1864 1933 ACYL CARRIER (ACP) 2.

FT DOMAIN 2910 2981 ACYL CARRIER (ACP) 3.

FT BINDING 820 820 PHOSPHOPANTETHEINE (BY SIMILARITY).

FT BINDING 1896 1896 PHOSPHOPANTETHEINE (BY SIMILARITY).

FT BINDING 2944 2944 PHOSPHOPANTETHEINE (BY SIMILARITY).

FT ACT_SITE 3502 3502 THIOESTERASE (BY SIMILARITY).

SQ SEQUENCE 3649 AA; 404079 MW; 6FD095704F858E6B CRC64;

Query Match 9.2%; Score 1484.5; DB 1; Length 3649;

Best Local Similarity 21.8%; Pred. No. 2.7e-77;

Matches 686; Conservative 424; Mismatches 1071; Indels 963; Gaps 104;

Qy 90 AKWEIVNQTALR-AFAFTSDSGTKSOVILKDSFVSMCWSSSSDPVVRDEAAAA-- 146

Db 923 AWQAARQTYPALRLRFDWAE--PVOIVDNDKPFWRPVDLSATADDAEQARVRELQ 979

Qy 147 -----ASGPRCNRFVLL---EDMTKKCOLVWTFSHALVDVTFOQVLSRVFAAYK 194

Db 980 ERDRTPEYDLAGGLRFLRYLIKQEDL---FSLIFCHHILDLGWSLPVLHDEVHRNYL 1035

Qy 195 HEKDTHRPETPSSDATDTSQSVSVSMSCEDNA-VSATHFWQTHLND----- 242

Db 1036 ----ALRAGQPIESDV-----DNAYVAAQRYWEAHRNDHAAYVWEOLGR 1075

Qy 243 -----LNASVFPPLS-----DHLMPNPTTTAEHR-----ITFPLSQKALS 279

Db 1076 IDERGDFAGLLNEKSRVRSVLDYDHVQ-----RHRTKLYLGADLGLKAGCAAD 1127

Qy 280 SAICRTALSTLLSRYTHS-----DEALFGA-VTQSLPFDKHYLDAGTYQTVLRVHCQS 334

Db 1128 QVTLHSLVQFVWHKVLHAIGGGNTTVGVTVSGRNLPGVDGSIENSAGLFIINTLPL----- 1181

Qy 335 NLRASDVMDAISYDDRLGLHAPFGLRDIR---NTGDNGSAA-----CDFQTV 379

Db 1182 -----IVDHDQAGQONVAEVRDIOAANVTMNSKISIVELGLRQSGMKRRRLFDTL 1231

Qy 380 LLVTGDSHVNNNGINGFLQQTITSSHPMPCNNRAI-----LHLCQMESSGAL 425

Db 1232 LVLENYPRL-----LDEEELAH-----QEARFEKAYDAKVDYPIAVVAREEGDEL 1279

Qy 426 LVA-----YYDHNVIDS-LQTTLLRLOQFGLHKLQCSPLDLSSMAEVNLMTEYDRAEIE 478

Db 1280 TVTLWYAGELFDEDTITDLVDARTL--FRQVTEIDIAP-----VRELDLISPMRARD 1332

Qy 479 SWNSQPLEV-QDTLLHHEMLKAVSHSPTKTAIQAWGDWTYSLELDNVSSRLAVHIKSLGL 537


```
QY 2579 QPVSDGQQRSAHMAPRTTEAILCDEFKVLGFQ-VGITDNFFDLGGHSLMATKLAVR 2637
| | : : | | : : | | : : | | : : | | : : | | : :
Db 3005 -----GNPQHISYNPPRDLVLEADLCRLWASALGTERCGIDDDDLFRIGGDSITAHIAAQ 3059
| | : : | | : : | | : : | | : : | | : : | | : :
QY 2638 IGHRLDTTYSKVDHPVLFQALALDNLV-----OSKTNEIYGGREMAEYSP 2686
| | : : | | : : | | : : | | : : | | : : | | : :
Db 3060 IHQIGRKVTVRDIFDHTTIRGIH---DNVWVKLVPHNVPOFAEQQTVLGD---APLLP 3113
| | : : | | : : | | : : | | : : | | : : | | : :
QY 2687 FQLLTEDPEEFMASEIKPOLELQELIIDIYPSTQMOKAFLEFDTHTTARPRPFVFFYIDFP 2746
| | : : | | : : | | : : | | : : | | : : | | : :
Db 3114 IQIWF-----SKSLQHPSHWNT-----FYL--- 3135
| | : : | | : : | | : : | | : : | | : : | | : :
QY 2747 STSEPDRAAGLIKACESLVNHLDFIRTVFAEASGELYQVVLSCLDLPQ--VIETED---- 2800
| | : : | | : : | | : : | | : : | | : : | | : :
Db 3136 RTPDLDTTTLSTAVAEQLYHDAFRMLRQIDGRTVQCFADDIS-PVQLRVLVNKVDGDS 3194
| | : : | | : : | | : : | | : : | | : : | | : :
QY 2801 -NINTATNEPDEF--AKEPVRGLCHPLIRETIKTKSMRVIMRISHALYDGLSLEHVVR 2857
| | : : | | : : | | : : | | : : | | : : | | : :
Db 3195 AADQQLQKQSDFDLEKGPICAAAYLHG-----EDSARVWFSVHHIIIDIYSWQILAR 3250
| | : : | | : : | | : : | | : : | | : : | | : :
QY 2858 KLHMLYN-----GRSLLPPHQSRYMQYTDGREGSHGFWRDVIQNT-----PMT--- 2902
| | : : | | : : | | : : | | : : | | : : | | : :
Db 3251 DLQILYEGGTGRKSSSVRWAEALQ-SYQCSASERAYWEGLLAQTAANISALPPVVTGTR 3309
| | : : | | : : | | : : | | : : | | : : | | : :
QY 2903 -----ILSDTVV-----DGNATCKALHLSKIVNIPSOVLRGSSNIIT 2941
| | : : | | : : | | : : | | : : | | : : | | : :
Db 3310 TRLARTWSDDRVTILLNEASNQASTODLLLAAGVGLALQQVTPGSPSMT 3359
| | : : | | : : | | : : | | : : | | : : | | : :
```

Search completed: May 30, 2003, 12:46:02
Job time : 139 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2003, 12:40:46 ; Search time 78 Seconds
(without alignments)
3856.469 Million cell updates/sec

Title: US-09-482-788-2

Perfect score: 16128

Sequence: 1 MEYLTAVDGRQDLPTPASF.....RVEHLLLEVSKEFEGNSSL 3129

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9720	60.3	3131	2 S39842	enniatin synthetase
2	5000	31.0	15281	2 S41309	cyclosporin synthetase
3	2275.5	14.1	4848	2 T30289	pristinamycin I synthetase
4	2084	12.9	6486	2 T31076	tyrocidine synthetase
5	1984	12.3	1997	2 T30874	virginiamycin S synthetase
6	1975	12.2	6359	2 T31679	bacitracin synthetase
7	1933	12.0	4450	2 JX0340	gramicidin S synthetase
8	1906.5	11.8	3670	2 T36249	CDA peptide synthetase
9	1896.5	11.8	2588	2 A12136	peptide synthetase
10	1882.5	11.7	5255	2 T31677	bacitracin synthetase
11	1876.5	11.6	9376	2 T14593	syringomycin synthetase
12	1867.5	11.6	5149	2 F83345	probable non-ribosomal peptide synthetase
13	1816	11.3	3587	2 T31075	tyrocidine synthetase
14	1810	11.2	7463	2 T36248	CDA peptide synthetase
15	1807.5	11.2	2555	2 C69681	peptide synthetase
16	1793.5	11.1	3587	2 I40486	surfactin synthetase
17	1781.5	11.0	2617	2 AE2136	peptide synthetase
18	1769.5	11.0	2560	1 I40457	peptide synthetase
19	1758	10.9	4452	1 YG8SG2	gramicidin S synthetase
20	1729.5	10.7	5369	2 T44807	mycosubtilin synthetase
21	1725	10.7	3603	1 D69681	peptide synthetase
22	1710	10.6	2607	2 T31678	bacitracin synthetase
23	1708.5	10.6	2448	2 S53999	pyoverdine synthetase
24	1677	10.4	4342	2 H83343	probable non-ribosomal peptide synthetase
25	1672	10.4	2611	2 T14591	actinomycin synthetase
26	1625.5	10.1	5232	2 A45086	HC-toxin synthetase
27	1612.5	10.0	2566	2 E98274	hypothetical protein
28	1611.5	10.0	2117	2 T36180	CDA peptide synthetase
29	1604	9.9	2561	1 I40456	peptide synthetase

30	1601	9.9	3588	2 I40485	surfactin synthetase
31	1538.5	9.5	4976	2 T14165	peptide synthetase
32	1532	9.5	2157	2 A13009	peptide synthetase
33	1515	9.4	3712	1 YGCEVC	alpha-aminoadipyl-pristinamycin I synthetase
34	1491	9.2	2591	2 T30288	pristinamycin I synthetase
35	1484.5	9.2	3649	1 S18268	delta-(L-alpha-aminoadipyl)-enniatin synthetase
36	1453.5	9.0	492	2 S53111	delta-(L-alpha-aminoadipyl)-enniatin synthetase
37	1422.5	8.8	3770	2 A40889	hypothetical protein
38	1402	8.7	2391	2 G89779	probable non-ribosomal peptide synthetase
39	1391.5	8.6	10797	2 T30192	saframycin Mx1 synthetase
40	1387	8.6	1317	2 B83346	hypothetical protein
41	1366.5	8.5	2605	2 T18552	alpha-aminoadipyl-pristinamycin I synthetase
42	1358	8.4	4077	2 T17484	hypothetical protein
43	1347.5	8.4	3791	1 YGPIV8	alpha-aminoadipyl-pristinamycin I synthetase
44	1346.5	8.3	3746	1 YGPIV3	lysobactin synthetase
45	1331	8.3	1575	2 T18545	lysobactin synthetase

ALIGNMENTS

RESULT 1

S39842

enniatin synthetase - fungus (Fusarium scirpi)

C:Species: Fusarium scirpi

C>Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 03-Nov-2000

C:Accession: S39842; S35906; S65363

R:Haese, A.

submitted to the EMBL Data Library, November 1992

A:Reference number: S39842

A:Accession: S39842

A:Molecule type: DNA

A:Residues: 1-3131 <HA>

A:Cross-references: EMBL:218755; NID:g2729; PIDN:CAA79245.1; PID:g2730.

R:Haese, A.; Schubert, M.; Herrmann, M.; Zocher, R.

Mol. Microbiol. 7, 905-914, 1993

A:Title: Molecular characterization of the enniatin synthetase gene encoding a multifunctional enzyme

A:Reference number: S35906; MUID:93247491; PMID:8483420

A:Accession: S35906

A>Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 499-1074; 1572-1988; 2423-2566 <HA2>

A:Cross-references: EMBL:218755

A:Experimental source: strain ETH 1536/J5

R:Pieper, R.; Haese, A.; Schroeder, W.; Zocher, R.

Eur. J. Biochem. 230, 119-126, 1995

A:Title: Arrangement of catalytic sites in the multifunctional enzyme enniatin synthetase

A:Reference number: S65363; MUID:95324513; PMID:7601090

A:Accession: S65363

A:Molecule type: protein

A:Residues: 2029-2048; 430-437; 1011-1020; 1021-1034; 1677-1695; 2294-2299 <PIE>

A:Experimental source: strain ETH 1536/J5

C:Genetics:

A:Gene: esyn1

C:Superfamily: acetate-CoA ligase homology; acyl carrier protein homology

C:Keywords: carrier protein; multifunctional enzyme; phosphopantetheine; phosphoprotease

F:531-985/Domain: acetate-CoA ligase homology <ACLI1>

F:1603-2100/Domain: acetate-CoA ligase homology <ACLI2>

F:2507-2574/Domain: acyl carrier protein homology <ACP1>

F:2601-2667/Domain: acyl carrier protein homology <ACP2>

F:1047,2538,2632/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 60.3%; Score 9720; DB 2; Length 3131;

Best Local Similarity 60.8%; Pred. No. 0;

Matches 1939; Conservative 443; Mismatches 691; Indels 114; Gaps 32;

QY 1 MEYLTAVDGRQDLPTPASFCSHGSDPLNSSYQLFHYLCIDSSRIEAIKPTFPQDLMI 60

Db 1 MSLLTSPDGGQDPALASKTLCT-----EQISRALGLGQDKIENIPGTFPFORDVI 49

QY 61 DCNALDKQSAIGHAVVDVPTDIDISRFALAWKEIVNOTPALRAFAFTSDSGSKTSOVLTKD 120

Db 50 DCAADDKQRAVGHAVFEIPKDKIDAAALAAAWKETVLHTPALTRCTCTTYSKSGDVLQVVLRD 109

[illegible]

RESULTS

RESULT:
S47309

341309 cyclosporin synthetase - cyclosporin fungus

C:Species: *Tolypocladium inflatum* (cyclosporin fungus)

C; Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text change 03-Nov-2000

C;Accession: S45487; S41309

R; Weber, G.; Schoergendorfer, K.; Schneider-Scherzer, E.; Leitner, E.

Curr. Genet. 26, 120-125, 1

A;Title: The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum*

A:Reference number: S45487; MUID:95094306; PMID:8001164
A:Accession: S45487
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-15281 <WEB>
A:Cross-references: EMBL:Z8383; NID:g440169; PIDN:CAA82227.1; PID:g440169
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993
A:Note: only a part of the translation is shown
A:Note: the source is designated as *Tolypocladium inflatum*
C:Genetics:

Query Match	31.08;	Score	5000;	DB 2;	Length	15281;
Best Local Similarity	32.4%;	Pred.	No. 9,1e-283;			
Matches 1233;	Conservative	440;	Mismatches	867;	Indels	1264;
					Gaps	66;
Qy	482	SQPLEVODTLIHHEMLKAVSHSPKTKTAIQAWDGDWTVSELDNVYSSRLAVIHKSJLGRQQ	541			
Db	11581	NQDPYPRDASVIDVFREQVASIPKSIADIVDASSQLTYTTELDERSSQLATWLRR-QVTVPE	11639			
Qy	542	AIIPVYFEKSKWVIASMLAVKGNATFLIDPNDPPARTAQVVTQT--RATVALTS----	595			
Db	11640	ELVGVLA PRSCETIIAEILGIKIANLAYLPIDVNAPAGRIETILSSLPENRLLILGSDTQA	11699			
Qy	596	-KLHRETVQKLVGRVVVDDELQSVS-ASDFESSLTKSQDLAVYFTSGSGDPKGIMI	653			
Db	11700	VKLHANSV-----RFRISDALVRSQPPPEELSTRPQAQSLAVYMTFTSGSTGVPGKGVV	11754			

Db 14131 LEQOSPIATMPLAEGIAQLRDAGALQMPKSDYPNALSILVDVFOQOAMASPSSTVAVTDS 14190
QY 2583 ----- 2582
Db 14191 SKLTYAELDRSLDQASYLRRQQLPAETWAVLAPRSCETIAFLAILKANLAYMPLDVN 14250
QY 2583 ----- 2582
Db 14251 TPSARMEAISSVFGRRLLVSGVRHADINVPNAKTMILSDVTGTDAIGTPEPLVVRP 14310
QY 2583 ----- 2582
Db 14311 SATSLAYVIFTSGTGKPGKGVWEHRAIMRLVKDSNVVTHPPATPMAHVNIADVSLF 14370
QY 2583 ----- 2582
Db 14371 EMCATLLNGGTLVCIDYLTLLDSTMLRETFEREQVRAAIFPALLRQCLVNMPPDAIGMLE 14430
QY 2583 ----- 2582
Db 14431 AVYVAGDRFHSRDARATQALAGRVRVYNAYGPTENAILSTIYNIDKHDPVNGVPIGSAYS 14490
QY 2583 ----- 2587
Db 14491 NSGAYVDRNQQLPPGVNGELVVTGEGVARGYTDASLDRFVTVTIDGQRQRAYRTGD 14550
QY 2588 ----- 2587
Db 14551 RVRYRPQGFIEFFGRDLQQAQKIRGHVRELGEVEHALLSENSVTDAAVVLRIMEEDPOL 14610
QY 2588 ----- 2587
Db 14611 VAFVTTDHEYRSGSNEEDPVATQAAGDMRKRSLPLPYVMVPSRVITLROMPLNANGK 14670
QY 2588 ----- 2629
Db 14671 VDRKLARAOQWPTASSGPPVHVAPRNETEAACIDFETILGVKVGITDNFFELGGHSL 14730
QY 2630 MATKLAVRIGHRLDPTVSKVDVDFHPVLFOALAIADNLV-----QSKTNETVGGREMAE 2683
Db 14731 LATKLAARLSRRMGLRISVKDLFDDPPVPSLAGLEQQGSGEDESSTVGIV----- 14783
QY 2684 YSPFOLLTEDEPEPMASEIKPOL-QEIIQDIYPTOMQAKFLDHTHTARPRFPVFPY 2742
Db 14784 --PFOLLPAEMSREIIQDVVPQIENGHSTPLDMYPATQTFIHLKRAKATGHPATPPLFS 14841
QY 2743 IDFPSTSPDAAGLLKACESLVNHLDIPTVFAEASGELYOVVLSCLDLPLOVIEDNI 2802
Db 14842 LDFPETA--DCRRLASAALVOHDIPTVFSRGGRFYQVLAHLDPVVEVIEQEL 14899
QY 2803 NTATNEFLDEFAKEPVRILGHPLIRPTIIKOT--KSMRVIMRISHALYDGLSLEHVVRKLHM 2861
Db 14900 DEVALALHEADKQQLRLGRAMLRITAILKRPCAKMRLVLRMSHSLYDGLSLEHIVNALHA 14959
QY 2862 LYNGSLPLPPHQFSYMOYTADGRSGHGHWKDVNTQPMITLSDDT--VVDGNDATCKAL 2920
Db 14960 LYSRKHQAQAPKFGFLYMHMASRRAGYFNWRSILQGSMTSLKRSVGALEAMTPSAGTW 15019
QY 2921 HLSKVTNIPSOVLRGSSNLTQATVFNACALVLSRESKDVFCGRVSGROGLPVEVQ 2980
Db 15020 QTSKIRIPPAALK---NGIQYATLFTAAVLSLLAKHTKSTDVFGVRSVQODLSINCO 15076
QY 2981 DIVGCTNAPVRAHI--ESSYNOLLHDIQOYLISLPHETIGFSDLKRNCTDMPPEAITN 3039
Db 15077 DIVGPCINEVPVRVIDEGDGMGLLRAIQOQYTSFRRHETLGLQEVKENCNTDWTATKE 15136
QY 3040 FSCCITYINFEPHPSQFEQORVEM----- 3064
Db 15137 FSCCIAFQNLNHPAEIEGOQIRLEGLPAKDQARQANGHPNGTNGTNGTNGANGT 15196
QY 3065 ----- 3102
Db 15197 NGTNGTNGHANGINGNSGVNCRSDNSVSAAGDAQAPVHDLDIVGPEPDGS--VKIGIGAS 15255

QY 3103 TQLFGKRKRVEHLLLEVSRTFEGN 3126
Db 15256 RQILGEKVVGSMLNLCETMLALS 15279

RESULT 3

T30289
pristinamycin I synthase 3 - Streptomyces pristinaespiralis
C:Species: Streptomyces pristinaespiralis
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 01-Dec-2000
C:Accession: T30289

R:De Crecy-Lagard, V.A.; Saurin, W.; Thibaut, D.; Gil, P.; Naudin, L.; Crouzet, J. B.
submitted to the EMBL Data Library, February 1997
A:Description: Streptogramin B biosynthesis in Streptomyces pristinaespiralis and Str
A:Reference number: Z20808

A:Accession: T30289
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4848 <DEC>
A:Cross-references: EMBL:Y11548; NID:e1025755; PID:e307539; PIDN:CAA72312.1
C:Genetics:

A:Gene: snbDE
C:Superfamily: acyl carrier protein homology; acetate-CoA ligase homology
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:511-951/Domain: acetate-CoA ligase homology <ACLI>
F:966-1034/Domain: acyl carrier protein homology <ACP1>
F:1563-2024/Domain: acetate-CoA ligase homology #status atypical <ACL2>
F:2399-2467/Domain: acyl carrier protein homology <ACP2>
F:2995-3427/Domain: acetate-CoA ligase homology <ACL3>
F:3441-3509/Domain: acyl carrier protein homology <ACP3>
F:4043-4492/Domain: acetate-CoA ligase homology <ACL4>
F:4507-4575/Domain: acyl carrier protein homology <ACP4>
F:998, 2431, 3473/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 14.1%; Score 2275.5; DB 2; Length 4848;
Best Local Similarity 26.7%; Pred. No. 9.8e-124;
Matches 877; Conservative 474; Mismatches 1280; Indels 659; Gaps 125;

QY 72 GHAVYDV-----PTDIDISREFAIANKEIYNQPALRAFAFTSDSGKTSOVLKDSFV 123
Db 31 GPDVTYVQVWFELRGPLEDEGURAAA--ALLURHPNLRAGFWQGVQGVVFPV 82
QY 124 FSWMCWSSSSSPDEV-----VRDEAAAAA-----GPRCNRFLV 157
Db 83 -----PNEVPLPWHTRDTFALGESDRERAEEAVVTADRAERFDPGAPPLIRFAL 131
QY 158 LEDMOTKKCOLVWTSFHALVDVTFQORVLSRVFAKYKHEKDT--HRPETESSDATDTS 215
Db 132 L-GLAADHHKLVLTTHLLLDGWSMPLLVRELTLYGQRGDAGMPVPTPYRAYLAHLAG 190
QY 216 QSVSVVSMSCEDNAVSAATHFWOTHLNDL-----NASVPHLSDHLMVP-----NPT 261
Db 191 R-----DQAAARAA--WRTALADLEPSLVAGAGAGAAAGSALPGQIWHIDA 239
QY 262 TTAEHRIPTPLSOKALSNAICR-----TALSILLSRYTHSDEALFGVTEQ- 308
Db 240 TTA-----GLTALARS--CNITLQHPGAKRLGTAARPQLGR---DDVVFQATVAHR 285
QY 309 --SLFPDKHYLDGTYQTVAPLRVHCOSNLRSDVMDAISYDDRLGLHAPGLDIRWT 366
Db 286 PPEIFGIESTI--GMFINTPLVRVRVPAETLGLLRVQREQAALIEHRHLSLTDIRST 343
QY 367 GNGSAACDFQVLL-----VTDGSHVNVNGINGFLQQTESHPNCRNALLHC 417
Db 344 TGTGEL---FDTVVVFENYPLDPVLRABEARGLRAGP--EVSATHY--PLSLAI----- 393
QY 418 QMESSGALLVAYDH--NVVDSLOTTRLQLQFGLIKCLQ-----SPLDSSMAEYN 467
Db 394 ---PGESLRFRLDHRGVDLDEAGARILLERLDTLLTDIAEHGADLPVGRDLSSAERH 449
QY 468 LMTYDRAIESWNSQPLEVQDTLTHHEMLKAVSHSPTKTAQWGDGWTSELONVSSR 527

Db 450 QV-----LEENDTGLPAEDATLALPEAQARTPDITALLVGRSLTYAELNARNR 502
QY 528 LAVHKSGLRAQAOAIIPYFEKSKWIASMLAVLKSNGAFTLIDNDPDPARTAQVVTOT 587
Db 503 LARHLVTLGAGPQO-IVAVKLESLDIYVALLAVLKTGAAYLPVDYTAERATFWMDDA 561
QY 588 RATVALTSKLRHTVOKLVGRCVVVDDELQSVASDDFSSLTKSODL-----AYVIFTS 642
Db 562 RPAVLTGE---DTGQDLSC---YDD-----TDLTDADRTAPLLPAHPAYVYTS 605
QY 643 GSTGDPKGMIEHRAFSKALCFGASIGINDSTRALQFGTHAFGACLELIMTTLINGCV 702
Db 606 GSTGTAKVMPGGAALNLLAMHURREIPGEAGAPVAQFTTIGEDVAAQELIATLWHLGKTL 665
QY 703 CIPSDDDRMNS---IPSFIRINRYN-----NMMAFSPYMGTFSPEDVPGLATLVLVGEQM 754
Db 666 AVPSQEVRRSAEQLAALWDEQHVSELYAPNLVIEALAEAAAEAGRTLPLALRHTAQAGEAL 725
QY 755 SSSVN-----AIWAPKLQLLNGYQSSSSICFASNMSTEP-----NNMGRAY-GAHSWV 803
Db 726 TLTRTVREFAAAVPGQLNHNHYPAE-THVMTGTALPEDPAWSEHAPLGRPVSGARVYV 784
QY 804 IDPNDINRLVPI--GAVGELVIESPGIARDYIVPPPEKSPFETD--IPSWYPANTFPDG 859
Db 785 LD-----SALRPVAPVTGELYLAGAGVSRGYNLRPVLTAEFRVADPYAFS-----PG 832
QY 860 AKLYRTGDLARYASDGSIVCLGRIDSOVKIRGORVELGAIETHLRQOMDDTLIVVEA-- 917
Db 833 ARMYRTGDLGRWNRGELEFAGRADHQVKIRGPRIEPEIEAAL-----TDLPAVARAAV 887
QY 918 TKRSQSANSTSLAFILIGSSYFGNRPDSDAHILDHDKATKAINKLEQVLPHRSPSYICM 977
Db 888 VREDRPGDKRLVAYAV-----PAGEGL-----DAA-AVRSLRALTLPDFMPPAAIVAL 935
QY 978 LELPRATGKIDRRIRIMGKOILDKOTQOAIQQAPALPVPFADT--AAKLHSIWQSL 1035
Db 936 DALPLTPNGKLDRAALPA-----PQSTAVQRTSPREEITLTLFAVLK---LPRV 985
QY 1036 GIDPATVNGATFELGSGNSITAIVKVNARSY-GMDLKVSNIYQHPTLAGISAVYKGD 1094
Db 986 GIDD-----NFFDLGGHSLATRLISRTVLGAEITLRLDFEAPTAVAGL-----GER 1033
QY 1095 LSY-----TLIPKSTHEGPVEQSYSGRLWFLDQDLGVSLWLYIPYAVMRGPNVDAL 1148
Db 1034 LQASGARPALTPAERPE-RLPLSHAORLWFLGRUDGNSTYNIPALRLSGTLDTAAL 1092
QY 1149 RRAALAEQRHETLRTTFEDQGVQO-----IVHEKLSSEMKVIDLCSGL 1195
Db 1093 RAALADLVARHETLRTYPSHDGEPYQHIVAPADATPALTVVH--TEHRLA-----1142
QY 1196 DPEVLNQOTTPFNLSAGWRATLRLGEDDHILTIYVHHIISDGSIDVLRDLNOL 1255
Db 1143 ---EQLAEATAPYDITLTELPRATLREGEDEHVLILLHHITAGDGWSLAPLTLRLAAA 1199
QY 1256 YSAALKSDPLSALTPLTOYSDFAKWK-----DQFIEQKOLNWKOLKDSPPA 1308
Db 1200 Y-AARDGNAP--GFAPLPQVADYTLWQRLGLGDADDTSLQAQALAHWQAL-BGRPA 1255
QY 1309 --KIPTDFARPALLSGDAGCVHVTIDGELYQSLRAFNCNEHNTTSFVLLAFAAAHYRLT 1366
Db 1256 HLELPCHDPRPATAVTHRGATVPFHIDAGLHEKLTALAKACDSSLFWLQAAFAALLTRHG 1315
QY 1367 AVEDAVIGTPIANRNPELEIDIGCVNTQCMKRNIDHHDHDTFTGLINQVATTTAAFENE 1426
Db 1316 AGTDIPVGSVAGRTDDALDDLVGFVNTLVLTDTSGDPTFRELVARVQRQDLAAATHQ 1375
QY 1427 DIPFERVVALQBSRDLSTPLAQILFAVHSO--KDLGRFEOGLSVVPVPSK-AYTRF 1483
Db 1376 DMPEKLVEBNP-ERSLARNPLFOVVLALQSMAPADLA---LPGLDVAAEPVRVGFARF 1431
QY 1484 DMFEHLFQETDS-----LKGSVNFADLFEKMTVENVRVFEFELNGLQSSRTPVSTL 1537
Db 1432 DLGLAVIEHTADGTTRAGIRGDWEYSTELFERCTVEALQSLRLVLLDAVDADDPERTIGSV 1491

QY 1538 PLTDGIVTLKLDVLNVKHVDYPRESSLADVFQTSAYPDSLAVVDSCLRLTYTELDRQ 1597
Db 1492 DLLDATERHRLVGLNTDGPPEERTLTALFQQAATPDATALVWGDRSLTYAELDAR 1551
QY 1598 SDILAGWLRRRSPMAETLVAFAPRSCETIVAFGVLKANLAYLPDVRSPSARVODILS 1657
Db 1552 ANRLARHLVALGAGPEQIVALQPLRSJLDLTVLAVMKSGAALPVPDPYPPERIAHMA 1611
QY 1658 GLSGPTVLILGHDTAPDTEVTNVEFRIDALNSNADGFEVIEHDSKTPSATS-LAYV 1716
Db 1612 D-ARPALV--DAIPDDTTLAAAYADSRLLTDA-----DRSAPLLPAHPAYV 1653
QY 1717 LYTSGTGRPGKGMIEHRIIVTITSGCIP-NYPSETRMAHMAITAFDGASYIYISALLF 1775
Db 1654 IYVTSGTGAPKGVVAHRSIAATVPAQAAAFGLGTHSRVLNPFASISFDAVWELTSALFT 1713
QY 1776 GRTLVCVDYMTLIDARALKDVFREHVNNAASHVTSQDVP-LRVPRRLSRTLMFFFLVV 1834
Db 1714 GAGLVADADDLLPGPSLARLVHHRHITLIALPPSALPALPDGALPPGTD-----LIVA 1767
QY 1835 TDSTAPDALDAQGLYOGVOCYNGYGTENGVMSTIYPIDSTESFINGV-PIGRALNNSGA 1893
Db 1768 GDATAFD--QAARFAPGRMWNAYGLTETVTCATM-----SEPATGAGPPIGRPAHARV 1821
QY 1894 YVVDPEQQLVGIVMGELVVTGDLGARYSDK-ALDENRFV-----HITVNDOTVKAYRTG 1948
Db 1822 YVLDERLRVPVPGVTGEMTVSGAGVARGYLHRPALTAQRFVDPDPYALLFGETGTRMYRTG 1881
QY 1949 DRVRYRIGIGLLEFFGMDTQKIRGNRIESABIEAALLRDSVRDAAVVLQNEOQAPE 2008
Db 1882 DLARLD-ADGRHFAHAGDQOVKIRGFIEPGIEITVLTAAHAPAAGAVIAREDTFGDKQ 1940
QY 2009 ILGFVVAD-HDISENDKGSANQVEGWQHFESGMSYDGEIDPS-TIGSDFKGWTSYMD 2066
Db 1941 LVAYLTRDTTHAAPPQDLGTGOVDRWQTYDS-----LYEAPGRAFGEDFSWNSSYT 1995
QY 2067 GSOIDDEHWEHUGETTRTLHDNRSLG----NVLEIGTSGGMILFNLDLSLESYVLEPS 2122
Db 1996 ABEIPLLEEREWRAATV-----DRVLALPRRVLEICCGTGLLSQVAPHTTEYRGTDLS 2050
QY 2123 RSAARFVNKATESIPSLAGKAKVQVGTATDIGOVDLHP--DLVVLNVIQVFPSEYL 2179
Db 2051 RSVVARLAHLGRPDADKVTVDARAA---HETDOLPKGRFTIYVLSVAQFFPDARYL 2107
QY 2180 AETADTLIHL--PNVQRIFFGVRSQATNEHFLLAARAIHTLGKNATKDDVRQKMALEDM 2237
Db 2108 AGILHRAEALLAPG-GTIFLGDILRLTLHAFRTAVELRRAGFPADPAAVRRRAVEQSLAT 2166
QY 2238 EBELLVEPAFFTSKDRFFGLVEHVEILPKNMEAVNELSAYRYAAVHVHVGSLGDELVL 2297
Db 2167 EXELLDDPDDFTDLVDRDPHLT-RADITLRTAAHNEMSRHRYDVTLLH-RG--GDS---P 2219
QY 2298 VEKDDWIDFOANQLNQLSLGDLKSSDAATMAVSKIPFEITAFERQVQVSLNSNIDEWOL 2357
Db 2220 AFAETVLRGT-----DLDTAGELARLVSGPRPVRTGIPN---ARLTGETRAAQA 2267
QY 2358 STIRSAEGDSSLSVP-----DIFRIAGEAGFRVEVSSARQSONGALDAVPHCC 2408
Db 2268 LERGSAAACARLEAPAGALDPHDVYAL-----GAVCTWGHG---DDTFDAYA 2314
QY 2409 SOG-----RTLNFPPDHLRGSDLLTNRLQKRNRIAEVREKLRLSLPSYMTPSNI 2463
Db 2315 GOGTAYRPRTRGQP-----LANDPARAEQDAALVELRDLVAQRLPAHMAPAAY 2364
QY 2464 VVLDKMLNANGKVDKELSRRAKVVYPKQTAAPLTPFIPISEVILCEATEVFCMKVD 2523
Db 2365 VILDRPLSANGKLDL-----DALPAP-----2386
QY 2524 ITDHFENLGGHSLLATKLISDRQLKVRITVKDVEDHPVADLASVIRQGLGLOQPVSD 2583
Db 2387 -----2386

Db 887 VIVEKDESGQNV-----LYAYLVN-----ERELTVAELEFGLRGLPSYMP 928
Qy 972 SFYICMLELPRATGKIDRRRLRIMGRDILDKOTQGAIVQOAPAPIPVFADTAAKLSIW 1031
Db 929 SFFIRLAEIPLTANGKVERKKL-----PRPAGAVVTGTAAAPQ-NEIEAKLAEIW 978
Qy 1032 VQSLGIDPATVNVGATFELGGNSITAIAKMY-NMARSVGMDLKVSNTIYQHTLAGISAVV 1090
Db 979 QQVILGI--SQVGIDHDPFDLGGHSLKAMTVFQVSKALEVELPVKALFEHPTVAELARFL 1036
Qy 1091 -KGDPISYTLI-PKSTHE-GPVEQSYSGRLWFLDQDVGSLWLYPIYAVMRGPNVDA 1147
Db 1037 SRSEKTEYTAIQVAAQETPV--SSAQKRWYILQQFEGNCISYNISGAILLEKLDYAR 1094
Qy 1148 LRRALAALEQRHETLRTTFEDQDQGVQIVHEKLSSEMKVIDICGSDLPDPFVLNORQTT 1207
Db 1095 FASAVOCLAERHEALRTSFHRIDGEVPQKVHEEVEPLFMLE--APEDAQAKIMREVR 1151
Qy 1208 PFNLSEAGNRYALLRLGEDDHILTIVNHHIISDGWSIDVLRDLNOLYSAALKDSKPL 1267
Db 1152 PFDLGVAPLMRTGLLKLGRHLFLLDMMHIIISDGVSSQILLREFAELYOGA-----1203
Qy 1268 SALTPLPIOXDFAKWQKDFIQ--EKOLNYYKKQKLDSSP-AKIPTDFARALLSGDA 1324
Db 1204 -DLOPLSLQKDFAAWONELFQYEAQKQEQHMLNTFPADEIPLNLPTDPRPSVOSFAG 1262
Qy 1325 GCYHVITIDGELYOSLRAFNCNEHTTSPVILLAAAFRAAHYRLTAVEDAVIGTPIANRNPE 1384
Db 1263 DLVFAAGKELLERLOQVASETGTTLYMILLAAVNVLLSKYTGOEDIIIVGTPVAGRSHAD 1322
Qy 1385 LEDIICGFVNTQCMRINIDHDTFGLINOVKATTTAAFNEDIPPERVVSALQPSRDL 1444
Db 1323 VENIMGIFVNTALRNQPASSKFTFAQFLQEVKONALAAAYDQDYPPEELVEKLAI-ORDI 1381
Qy 1445 SSTPLAQIFAVHSQDGLGRKFQGLSEVPVPSKAY-----TREDMEFHLFOETSLKG 1498
Db 1382 SRNPLEDTLSLENANO-----QSLAIAELTASPYELFNKISKFDLALNASESPADIQF 1435
Qy 1499 SVNFADLFKMETVENVVRVFFELRLNGLOSSRTPVSILPLTDGIVTLEKLDVL---NVK 1555
Db 1436 QLTFAATKFKKETVERMARHYLEILR---WISQPTASLADIDMTAEAKRTLLLVNDT 1492
Qy 1556 HDVYPRESSLADVFQVSAYPSLAVVDSSCRILTVELDQSDILAGWLRSMFAETL 1615
Db 1493 FVERTAALHQLVEQQAARTPEVAVVEEYALTRELNARANQALRLRSHGTGPDTL 1552
Qy 1616 VAYFAPRSCETIIVAFFGLKANLAYPLDVRSPSARVQDILSLSGTIVLIGH--DTAP 1673
Db 1553 IGIWDRSPGMVGVMLAVLKAGAYTIDPSYPERIQYMLSDSQAPILLTQRHLQELAA 1612
Qy 1674 PDEVTNVEFVRTRDALNSNADGFEVIEHDSKPSATSILAYVLYTSGSTGRPKGYMIEH 1733
Db 1613 YQGEIIVDDEAYTGA-DTNLDNV-----AGKDDLAYVIYTSGSTGNPKGVNISH 1662
Qy 1734 RVITRVTSOCINYP--SETRMAMHATIAFGASVEIYSALLFGRTLVC-----VD 1783
Db 1663 QALCNHML-WMRTFTPLTTEDAVLQKTPFSFASVWVEFVPLITGGOLVLAKPDGHRDIA 1721
Qy 1784 YMTTLARALKQVFFREHVNAAASHVTSQSDQVPLVRPRLSRRLMFFFLVVTDTAPDAL 1843
Db 1722 YMTRL-----IRDEKITTLQWVPSSLDDLVMTPDQWSACTSLQVFCGGGALTALPV 1772
Qy 1844 DAQGLYQGVQCYNGYPTENGVMSTIYPIDSTESFNGVPIGRALNSGAYVVDPEQQLV 1903
Db 1773 SRFYETQQAOLINLYGTETTTDATYWPQROQY--SAIPIGKPIDNVRLYVNVNASNQLQ 1831
Qy 1904 GIGVMGELVVTGDLARGY-SOKALDENREHVHTVNDQTVK---AYRTGDRVRYRIGDG 1958
Db 1832 PVGAGELMCIAGDLARGYWQREULTKASV-----DNFPFGTMYRTGDMVRY-LPDG 1885
Qy 1959 LIEFFGRMTQKIRGNRIESABIEAALLRDSRDAAVVLQONEDQAEIILGFVVADHD 2018
Db 1886 HIEYLGRIDHVKIRGHIELGEIATLLOHEAVKAVVMVNRQDGKQNSLAYVVAEQD 1945

Qy 2019 -----HSENDKGOSA-----N 2029
Db 1946 IOTAEALTYLSATLPAYWPSAFVLEQLPLSANGKYDRKALPOPEDAASAAVYVAPRN 2005
Qy 2030 QVE-----GQWQDH-----2038
Db 2006 EWEAKLAAIWESVLGVEPIGVHDHFFELGGHSLKAMHVISLLQORSFOVDPLKVLFSPT 2065
Qy 2039 -----ESCMYSDI-----G 2047
Db 2066 IAGLAPLVAARAGTYTAIPVEKQYYPVSAQKRMFILOOMEGAGISYMPGMYLDG 2125
Qy 2048 EIDPSTIGSFKGWTSMYDGSQIDFDEMHWLAGEETTRTLHDNRSLG-----2093
Db 2126 KLQTERLOQALKSLVQRHESLRTSF---HSVOGETVQRVDDVDLALISFGEATEAETROI 2182
Qy 2094 ---NVLETGTG---SGMI-----LFLNLSR-----LESYVGL-----2119
Db 2183 AEOFIOPFDLGTAPLLRAGLIKLAPERHLFMDLHHIWDGVSIGILLIEFAQLYHGEEL 2242
Qy 2120 -----EPSSAAAFVNKATESIP-----SLAGKAK 2144
Db 2243 PALRIQYKOPAKWQDMWFQTEFAEQAYWLNTFTGEIPVNLPTDPRPSVKSFAGDRF 2302
Qy 2145 VOVGATADI-GQVDDDLHPD-----LVVLNSVIOYFSSSEYLAETADTLIHLNVQR 2194
Db 2303 V-FSGGTALPKOLHQLAQETGTTLYMWLLAAYNVLLSKYSRQE-----DIIIGCAPTAGR 2355
Qy 2195 -----IFRGD---VRSQ---ATNEHFLAARAIHTLGNKATKDDVRQKMAELBDE 2238
Db 2356 SHAETESIVGMFVNTLALRNEPAGGKTRDFLAEVKINTLGAFEHQDYPLDELVDKIDMQ 2415
Qy 2239 EELLVPEAPRTSL---KDRFPGVLEHVEILP-----2266
Db 2416 ROLSRNPLFTVFIQNMEOKPFEMEBOLITTPYSAEVKQAKFDLSLEAYEENABIIFSLD 2475
Qy 2267 -----2266
Db 2476 YSTKLSFRETIERKIATHFIQILRAVIAEPEMLSEITMLTAEAKORLLVDFNGAKHDPQ 2535
Qy 2267 -KNWEAVNELSAYR--YAAVVHVRG-----SIGDELVL-----VEKDWDIDPOAN 2309
Db 2536 NKTLOALFEEQAEKSPQATAVEISGQPSLSYQELNERANQLAATLRERGVDQDPQVIGMAN 2595
Qy 2310 QLNOKSIGDL--LKSSDAAIMAVSKIPFEITAF-----PROVVASLNSNT---2352
Db 2596 RSVEMVVGILAILKAGAYVIDPEYPERVAVMLTDCQARLVLTOKHIGAKLGSSVTA 2655
Qy 2353 ---DEWOLSTIRSAEGDSSLS-VPDIFRIAGEAG-----F 2384
Db 2656 CLYLDDESNGVHRSNLQPIINTASDLAYIYTSGTGKPKGVMMVHRGIVNNVLWKA 2715
Qy 2385 RVEVSSARQWSQALDAV---FHHCCSQGRTLVNFTDHLRG-----SDLL 2429
Db 2716 QMKVGRSLSLSFAFADFVLSFTFVLSCATVV-LAEDEEAKDPVSLKLLIAASRCTL 2774
Qy 2430 TN-----RPLQ--RLQNRRIAEVRERLRSLLPSVMI-----2459
Db 2775 TGVPSLFQALLECSTPADIRPLQTVILGGEKITAQLVEKCKQLNPLDIVINEYGPTESSV 2834
Qy 2460 -----PSNIVVLDK-----MPLNANGK---VDRKELSR---RAKV 2488
Db 2835 VATWORLAGPDAAITIGRPANTSLYVNOYHQLQPIGVVGEICIGRGLARGYWNKPAL 2894
Qy 2489 VPQKQTAAPLPT-----FPISEVEVI-LCEATEVEFGMKVDTID-----2526
Db 2895 TEKEFVSHPFAAGERMYKTGDLGKWLDPGTIEYIGRIDEQVKVGRYRIEGETESALLAA 2954
Qy 2527 -----HFNLCGHSLA-----TKLSRISDORLKVITVKVDFHDPVAD 2566
Db 2955 EKILTAVVVVYEDQLGQSALAAAYFTADEQLDVTKLWSHLSKRLPSTY-----IPAHFVOLD 3010

[illegible]

Db 2476 HTTENOMEYBKDTISRLNHQKLERLEQIIKHCTOQTESERTPSDYGDTNISLAE----- 2529

Qy 2700 ASEIKPOLEQEIQDIYPTSMQKAFLE----DHHT-----ARPRFPVFIIDPPSTSEP 2751

Db 2530 LEEIKG--KYRSALEKIYPLANQKMLFAIEDHTSDAYFOQIVMDIEGYVD-PAILEA 2586

Qy 2752 DAAGLIACSLVNLHDIPTVFAEASGELYQVVLSCDLDPQIVETEDNI-----NTAT 2806

Db 2587 SFNDIMKRHEIL-----RASYE-YEIV-----EPRQIIENRSIDETIFYNIAK 2629

Qy 2807 NE-----FLDEFAPKPEVRLGHPLIRFTIHK-----QTKSMRVIMRISHALVDGLSLEH 2854

Db 2630 SSAQQOQEMFIERLLNEDRKKGFDLSKVDLMRAYLLKTAERSYRLVWSHHILLDGCWCLGI 2689

Qy 2855 VYRKLHMLY----NGRS--LLPHPHQFSRYMOY-TADGRESHGFWRDVI-----ONTPTMTI 2903

Db 2690 IMRELFIYENRMNGKASPLKETPKPSDYIKWLERQDQDEARQVWREYLKGYEEQAOLPT 2749

Qy 2904 LSDTTVDGNDATCKALHLSKIYNIPOVLGSSNIITQATVFNACALVLSRESDSKOV 2963

Db 2750 LTRKKSSRYDRREKVIHLSKQLTKOLKEL-AAKNSVTLTHTVIQTWGLMLTRYTKIDDV 2808

Qy 2964 VFRIGVSGROGLPVEYDIIYGPCTNAVVPRAHI-ESSDYNOLLHDIDQVLLSLPH---- 3018

Db 2809 VFTGVVSGREANVDGIEDMIGLFINTIPTIRRENOARFNDCLQKQVEDAIQSNRYNYM 2868

Qy 3019 --ETIGSDLKRNCTDMPAITNFSCITVYHNFEYHPESQFOQRVEMGV----- 3066

Db 2869 LAEQVALLSLKKDLIDH-----ILVFNENY-ADEODFEESQMKTKFKVNEISAEEQ 2918

Qy 3067 LTKF-----VNIEMDEPLVDLAGEVEPDGAGLKVTYAKTOLFGRK--RVE-3112

Db 2919 SITAFMSVTPGBELTLVLTYDNGVYDRDIINNIE---GHIKRVAEQVTANENRKTAEID 2975

Qy 3113 HLLIEVSKT 3121

Db 2976 MLAEERKT 2984

RESULT 7

JX0340

gramicidin S synthase 2 - Bacillus brevis

N;Alternate names: gene Grs2 protein

C;Species: Bacillus brevis

C;Date: 22-Apr-1995 #sequence_revision 26-May-1995 #text_change 03-Nov-2000

C;Accession: JX0340

R;Saito, F.; Hori, K.; Kanda, M.; Kurotsu, T.; Saito, Y.

J. Biochem. 116, 357-367, 1994

A;Title: Entire nucleotide sequence for Bacillus brevis Nagano Grs2 gene encoding gramicidin S synthase

A;Reference number: JX0340; MUID:95122465; PMID:7822255

A;Accession: JX0340

A;Molecule type: DNA

A;Residues: 1-4450 <SAI>

A;Cross-references: DDBJ:D29676; NID:g473695; PID:g511490

C;Comment: This enzyme catalyzes activation and combination of four constituent amino acids to form gramicidin S. The enzyme contains 4 domains with a mean of 1042 amino acid residues containing 11 alpha-helices and 10 beta-sheets.

C;Superfamily: alpha-aminoacyl-tRNA synthetase

C;Keywords: carrier protein; phosphopantetheine; phosphoprotein

F;514-956/Domain: acetate-CoA ligase homology <ACLI>

F;974-1042/Domain: acyl carrier protein homology <ACPL>

F;1552-1991/Domain: acetate-CoA ligase homology <ACL2>

F;2009-2077/Domain: acyl carrier protein homology <ACP2>

F;2588-3036/Domain: acetate-CoA ligase homology <ACL3>

F;3054-3122/Domain: acyl carrier protein homology <ACP3>

F;3634-4074/Domain: acetate-CoA ligase homology <ACL4>

F;4092-4160/Domain: acyl carrier protein homology <ACP4>

F;1006,2041,3086,4124/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 12.0%; Score 1933; DB 2; Length 4450;

Best Local Similarity 22.0%; Pred. No. 1.le-103;

Matches 845; Conservative 587; Mismatches 1306; Indels 1098; Gaps 137;

Qy 67 KQSAIGHAVYDVPDIDISRFALAWKEIVN---QTPALRAFAPATSDSGKTS----- 114

Db 19 QEGMLFHALDKOKNAHLVQMSIAIEGIVDELLSESLNILLIDRYDVFRTFLHEKIKQP 78

Qy 115 -QVILKDSFVFSWMCWSSSSSPDEVVDEAAAA---ASGPRCNRFLVED-----MQT 163

Db 79 LOWVLKERPV--QLQFKDISLDEEKREQAIEQKYQDGETV--FDLTROPLMRVAIFOT 134

Qy 164 KKC--QLVWTFSHALVDVTFQQRVLSRVFAAYKHEKDTHRPETPESDSDATDTSQSISVV 221

Db 135 GKVNQMINSEFHLLMDGWCFFIIFNDLFNIYLSLKE-----KKPLQLEAVQPKQFKWL 190

Qy 222 SMCEDNAVSAATHFWQTHLDNLNASVPHLSHDHLMVNPNTT--TAEHRTTFP----- 271

Db 191 E---KDKQKQALRYWKEHLNMDQSV--TLPKKKAAINNTTYEPAQFPAFADKVLTOQLL 245

Qy 272 --LSQKALSNAICRTALSILLRYTHSDALFAGVTEQSLPDKHYLAD--GTIQTIVAP 327

Db 246 RIANQSOVTLNIVFQTIWIGIVLOKYNSTNDVYGSVV--SGRPSIEISGIEKMWGLFINTLP 304

Qy 328 LRVHCQSNLRASDVMDAISS---YDRLGHLAPGLDIRNTGDNLSAACDFQVL--- 380

Db 305 LRITQKQDSFIELVKTIVHONVLFSSQHEY---FPLYEQNHTELKONLIDHINVIENYP 361

Qy 381 LVTDGSHVNNINGFLQO-----ITESHFMPCNNRALLLHCQMESSGALL---VAYYDHN 433

Db 362 LVBELOK-----NSIMQKVGFTVRDVKMFETNYDMTVMLPRDEISVRLDYNAAYVDID 416

Qy 434 VIDSLOTTRLLQOFGHLIK--CL-QSPDLSSMAEYNLATEYDRAE--IESWNSQPLEVQ 488

Db 417 FIKRIE-----GHMKEVALCVANNPHVL--VQDVPLLTQKOEKHLVVELHDSITEYP 466

Qy 489 DTLIHHEMLKAVSHSPTKTAIQAWDGDWYSELNDVSSRLAVHITSGLRAQAAIIPYVF 548

Db 467 DKTIHQLETFQVEKTPHVAVVFEDKVTYRELHERSNQLARLREKGVK--KESIIIGMM 525

Qy 549 ESKWVIASMLAVLKSGNAFTLIDPNDPPARTAQVVTQTRATVALTSKLRHRETVQKLVGR 608

Db 526 ERSVEMIVGILGILKAGGAEVPIPEYKPERIGYMLDSVR--LVLTQR--HLKOKFAFTE 582

Qy 609 CVVDDDELQSVASDODFSLSLTKSQDLAVYVFTSGTGDPKGIMIEHRAFSSCALKRFGAS 668

Db 583 TIVIEDPSI-SHELTEIDYINSEDLFYIYISGTTGKPKGVMLHKNIVNLLHFTFEK 641

Qy 669 LGINSTRALQFTHAFGACLLIMTTLINGG--CVCIPSDDDR--MNSIPSFIRNYNNWM 726

Db 642 TNINFSDKVLQYTTCSFDVCYQEIFSTLLSGGQLYLIRKETQRDVEQLFDLVKRENIE-V 700

Qy 727 MATP-----SYMGTFSPEVDYVGLATLVLGEOM---SSSVNAIWAQKLOLLNGY 772

Db 701 LSPFVAFLEKPIFNREIFNRF-PTCVKHIIT---AGEQLVNVNEFKRYLHEHNVHLNHY 756

Qy 773 GQSESSI-CFASNMSTEPNNM---GRAVGAHSHWIDPNDINRLVPICAGVELVIESPGI 828

Db 757 GPSETHVITYTINPAEIPPELPICKPI-SNTWIYILDOEQQLQPOQIVGELYISCANV 815

Qy 829 ARDYIVPPPEKSPFTTDPISWPANTFPDGAKLRTGLARVADSGSIVCLGRIDSQVK 888

Db 816 GRGYLNNQELTAKEKFFAD-----PFRPNERMYRTGLARMLPDGNIIEFLGRADHGVK 867

Qy 889 IRGQVELGAIETHLRQOMPDDLTIVVEATKRSQSANSTSLIAFLIGSSYFGNRPSDAHI 948

Db 868 IRGRIELGEIEAQLNCKGVKQAVVID---KADDKGGKYLCAIVV-----MEVEV 915

Qy 949 LDHDATAKINIKLEQVLRHSIPSFYICMLELPELRTATGKIDRRRLRMGDKILDQKQGA 1008

Db 916 NDSE---LREYLGKALPDYMIPEFVPLDOLPITNGKIDRSL-----PNLEGI 962

Qy 1009 IVQQAAPATPVFADTAALKHLSIWQSLGIDPATVNVGATFFELGGSNITAIKMN--MARS 1067

Db 963 VNTNAKYVVPT--NELEEKLAKIWEVLGI--SQIGIQDNFFSLGHSKLAITLISRMKE 1019

Qy 1068 VGMDLKVSNIYQHTPLAGISAVVKG--DPLSYTLIPKSTHSGPV-EQSY-----SQGLRWF 1120

Db 1020 CNDIPLRLLFEAPTIOBISNYINGAKKESYVAIQ-----PVPEOEYYPVSSVQKRMFI 1073
QY 1121 LDOLDVGLSWLPIYAVYMRGPVNDALRRALALEQRHETLRTTFEDODGVQIVHEK 1180
Db 1074 LNEFDRSTAYNLFGVWMLDGLKNVRLQLEAAVKLVERHEALRTSFHSINGEPQVRVHQN 1133
QY 1181 LSEEMKYIDLCSOLDPEVLVNOEQTPFPNLSSEAGWRATLLRLGEDDHILTIYVMHHIS 1240
Db 1134 VELQIAYSE---STEDQVERIIAEBFMOQFALEVPALLRVLGLVLEAERHLFTMDMHHIS 1190
QY 1241 DGWSDVLRDLNOLYSALAKDSKPLSALPLPIQYSDPAKW-----OKDOFTEQKOL 1295
Db 1191 DGVSQIMQIADLYKE-----KELPTLGIQYKDFTVWNNRLQSDVI---EKQE 1238
QY 1296 NWWKQLKDDSPA-KIPTDFARPALLSGDAGCVHVTIDGELYQSLRAFECNEHNTSFVVL 1354
Db 1239 AYWLNVFTEETPVNLFPDYPRPTIOSFGDGRFTSTGKQLMDDLYKVATETGTYLWVL 1298
QY 1355 LAAPRAAHYRLTAVEDAVIGTPIANRNPELEDIICFVNTQCMRINIDHDHDTGTLNQ 1414
Db 1299 LAAYNVLSKYSQODDIWVGTPAGRSHADVENMLGMPVNTLAIRSLNNEDTFKDFLAN 1358
QY 1415 VKATTAAAFENEDIPFERVWSALOPG-SROLSSTPLAOLIFAVHSQKDLGRKFGLESV 1473
Db 1359 VKOTALHAYENPDYFDTLVEKL--GIQRDLSRNPFDFTMEVLQN-TDRKSFEVEQITIT 1415
QY 1474 P-VPSKYATREDMEFHLPOETDLSKGSVNADELFKMETVENVVRFEILLRNGLOSRST 1532
Db 1416 PYVNSRHSKDFLEWSEONEILLCEYCTKFTDKTVERMAGHFIQIILHAIVGN--- 1472
QY 1533 PVSILPTDGTVITLEKLDVL---NVKHVDYPRESSLADVFQTVQSAYPDLSLAVDSSCRL 1589
Db 1473 PTIIISEILSEEEKOHILFEENDTKTTPHMOTIOGLFEEQVEKTPDHVAVGWKQDAL 1532
QY 1590 TYTELDRSDILAGLWRRSPMAETLVAFAPRSCETTIVAFGVKLANLAYLPDLVRSPS 1649
Db 1533 TYRELNERANQVARYLROKGVQOPDNIVGLLVERSPEMLVGIMGLKAGGAYLPDPEYPA 1592
QY 1650 ARVODILSGLSGPTVLIGHDTAPPDIEVTNVEF---VRIDALNDSNADGFEVIEHST 1706
Db 1593 DRISYMIOD-CGVRLM---TQHLLSLVHDEFDCVILDEDSLYKGSNSLAPVN--- 1643
QY 1707 KPSATSLAYLYTSGSTGRKPGVMEHRIIIVTSTGCPINYPSETRMAHMATIAFDGAS 1766
Db 1644 --QAGDLAYIMVTSSTGCKPGVMVHRNVRILVKNTNVQVREDDRIIQTGAIGEDALT 1701
QY 1767 YEIYSALLFGRTLVCVDYMTLIDARALKDVFREHVNAAASHVTSSQDVPRLVPRRLSRT 1826
Db 1702 FEVFGSLRHGAELPVTKDVLDDAEKHLKFL-----QANQITIMWLTSPLF--NQLSQG 1753
QY 1827 LMFFF-----LVVTDSPADAL-DAOGLYOGVOCYNGYPTENGVMSTIYPIDSTESI 1879
Db 1754 TEEMFAGRLSLVGGDALSPKHNNVKKCPNLWNGYGTENTTFTSCFLID--KEYD 1811
QY 1880 NGVPTGRALNNAGYVVDPEOOLVGIGVMGELWLTGDLGARYSDK-ALDENRFVHITVN 1938
Db 1812 DNIPIGKAISNTVYIMDRYQOLQPVGPVGCGLVCGDGVARGYMNQNPATEEKFVPNPFA 1871
QY 1939 DOTVKAYTGRVRYRIGDGLIEFGRMDTQFKIRGNRIESAETEAALLRSSVDDAAV 1998
Db 1872 PGE-RMYRTGDLAR-LPDGTIEYLGRIDQOQVKIRGYRIEFGPEIETLLVKKKKYESVIM 1929
QY 1999 LQONEDOAPETILGFVVADHD-----HSENDK--- 2024
Db 1930 VVEDNNGOKALCAYVPEEVTVSELREYIAKELPVYVMPAYFVQIEQMPLTQNGKVNRS 1989
QY 2025 -----GOSANQVE-----GWQDHF----- 2038
Db 1990 ALPKPDGEGTATEVAVPSSDIEMKLAETIHNVLGVNKGVLNDNFFELGGHSLRAMTIS 2049
QY 2039 -----ESGMYSDIGELDPT-----IG 2055
Db 2037 -----PEPDGSIISIGTIEVAPRTMLEKLEIWKDVLGLQVRGIIHDDFFTIGHSLKA 3089
Db 2050 QVHKFEFDELPLKVLFETPTTISALAQYIADGKGMYLAIQVPVTPODYYPVSSAQKRMVIL 2109
QY 2056 SDFKWTSMYDGSQIDFDE-----MHEWL-----GETRTRLHONR 2090
Db 2110 YEFEGAGITYNPNVPMFIEGKLDYORFEYATKSLINRHEALRTSFYSLNGEPVQVRHONV 2169
QY 2091 SL-----GNVLEIG-----TSGSMILFNLDSR----- 2112
Db 2170 ELQIAYSEAKEDETEBQIVESFQPPDELEAPALRVGLVKLASDRHLFMDMHHISDGV 2229
QY 2113 -----LESYVGLEPS-----RSAAA-----FVNKATESIPSL-- 2139
Db 2230 MQLITKEIADLYKGLAEHLHIQYKDFAVQWNEWFQSAALEKQKTYWLTFAEDIPVNL 2289
QY 2140 -----AGRAK 2144
Db 2290 STDYPRPTIOSFEGDIVTFSAGKQLAEELKRLATETGTYLMLLLAAYNVLLHKGSGQEE 2349
QY 2145 VOVGTA-----TDIGQVDDLHPD-----LVLVNSVIOYFPSEYLA 2180
Db 2350 IWVGTPAGRSHADVENIVGMFVNTLALKNTPIAVRTFHEFLLEKVNQNALEAFENQDYPF 2409
QY 2181 E-----IADTLHLPNV-QRIFFG-----DVR 2201
Db 2410 ENLIEKLOVRDLRSRNPFDTFMFSLSNIDEQVEIGIEGLSFSPYEMOYWIAKFDSFIDL 2469
QY 2202 SOATNEHF-----LAARAIHTLGNATKDDVROKMAEL-----EDMEBEL 2241
Db 2470 EKODDIOFYFNCTNLFKKEITERLATHFMHILQEIIVINPEI--KLCEINMLSEEQORV 2527
QY 2242 LVE-----PAFFTSLKDRFPLGVHVEIILPKNMEAYNELSAYRY----- 2280
Db 2528 LYDFNGTDATVATNKIFHELFEQVEKTPDHIAVIDEREKLSYQELNAKANLARVLROK 2587
QY 2281 -----AAVHVVRG 2288
Db 2588 GVQPNMVGIMVDRSLDMIVGLKAGGAYVPIIDIDYPOERISYMMEDSGAALLTQO 2647
QY 2289 SLGDEL-----VLPVEKDDWIDFOANQLN----- 2312
Db 2648 KLTQOIAFGDILYLDQEEWLEHEASLEPIARPQDIAYIIVTSGTTPKPGVMIEHOSY 2707
QY 2313 -----OKSLGDL-----LKSSDAAIM 2328
Db 2708 VNVAMAKDAYRLDTFPVRLQLMASFAFDVSAGDFARALLTGGLQILVCPNEVKMPASLY 2767
QY 2329 AVSKIPFEITAFER--QVVASLNSNIDEMQL--STIRSAEGDSSLSVPDIFRIAGEAGF 2384
Db 2768 AIIR-KYDITITEATPALVPIPLMEYIEOKLDISOLQILVIGSDSCMEDFTLVSRRFGS 2826
QY 2385 RVEV-----SSARQWSQNGALD-----AVPH 2405
Db 2827 TIRIVSYGVTEACIDSSYEQPLSSLHVTGTVPIGKPYANKMYIMNQYLQIQPVGVI 2886
QY 2406 HCCSQR-----TLVNFTPDHHLRGSDDLTLNRPLOR----- 2436
Db 2887 ELCIGGAGVARGYLNRPDLTAEKFPVPEKLYRTGDLARMWPDGPNVEFLGRNDHQV 2946
QY 2437 -LQNRRIAL-----EVRERLRS 2452
Db 2947 KIRGIRIEGETEAOLRKHDSIKEATVIAREDHMKKEYLCAYMVTEGEVNVNLAELRAYLAT 3006
QY 2453 LPLSNMIESNIVVLDMPLNANGKVDKRLSRAKVRVVKQQAAPLPTFPISEVEVILCE 2512
Db 3007 DLPAAMIDPSYVLEAMPLTANGKIDKRSI----- 3036
QY 2513 EATEVFGMKVDITDHFNLGGHSLLATKLISRIDORLKVRTVKDVFDPHFADLASVIR 2572
Db 3037 ----- 3036
QY 2573 QGLGLQPVSDCQODRSNHMAPRTETEAILCDEFKVLGFQ-VGITDNFFDLGGHSLMA 2631
Db 3037 -----PEPDGSIISIGTIEVAPRTMLEKLEIWKDVLGLQVRGIIHDDFFTIGHSLKA 3089

Db 1004 FETPTAGLSALDTSGLTRLTALTARPPEIPISYAQORLWFLHOLEGPTATNIPITL 1063
QY 1138 RMRFVNVDAURLAALAEORHETLRTTFEDQGVGOIV--HEKLSSEMKVIDLGGSDL 1195
Db 1064 RLCTGLTDALQSFANDLLARHESLRTYTTEDDGAQOIVLPVAVATPPFAVVDVAEDV 1123
QY 1196 DPFEVLNQEQTTFFNLSEAGWRATLLRLGEDDHLITVMHHIISDGWSIDVLRDLNQL 1255
Db 1124 --AERVAEAAAHAFDGAIEIPVRLFRVSRHVVLLLVLHHIASDAWSRGPLAQDLTAA 1181
QY 1256 YSAALKDSDPLSALTPLPIQYSDFAKWOKDOFEOE-----KOLNYWKKOLK--SSP 1307
Db 1182 YTARCADD---APAWQPLPVQYADYALWQOILGDDTDPDTLAGRQLAYWKKOQLAGLPEQ 1238
QY 1308 AKIPTFARPALSGDAGCVHTIDGELYQSLRAFNCHEHTTSFVLLAFAFRAAHYRLTA 1367
Db 1239 LDLPTRPRATADHTGDRVTFTVPADLHTLRLTELARETNTAFWVIOQAALALTRHGA 1298
QY 1368 VEDAVIGTPIANRNRPELIEDIGCFVNTQCMRINIDHHDTEGLINQVKATTTAAFENED 1427
Db 1299 GEDIPIGTPVAGRTDDATDHLIGFEVNTLVLRDTSGNPTFRDLTLRVRODTLTAYTHQD 1358
QY 1428 IPFRVYSALQPGSRDLSSSTLAQLIFAVHQSOKDLGREKFGLESVPVPSKAYT--RFDME 1486
Db 1359 LPFERLVEALNP--TRSLTHHPLFQTMLTLLHNTQCTKTDRFAGLAAEVVASESVSAREFDS 1417
QY 1487 FHLFOETDS-----LKGSVNFADELFKMETVENNVVRFELRLNGLQSSRTPVSILPLT 1540
Db 1418 FALAHEFGADHSCAGMSGVITYSTALFDRATVRDLADRLVLLAAVAHGRSGVQOLEIM 1477
QY 1541 DGIWTEKLDVNLKHHVDYPRESSLADVFQTSAYPDSLAVPSSCLTYTDELDRQSDI 1600
Db 1478 DAAERRLVQEWNDTAAB--PPAASVTGLFERQARRSPGATALEFCEVRSVLAELNARANR 1536
QY 1601 LAGLRLRSMFAETLVAVFAPRSCTEIVAFPGVLKANLAYLPDLYRSPSARVODTLISLS 1660
Db 1537 LARHLVGAAGAPGRFVAVLPRSAELVVTLIAVLKSGAAYLPIDPHYPADRVVEYMLAD--A 1595
QY 1661 GPTVLIGHDTAPPDIEVTNVEFVRIRDALNDSNADGPEVIEHDSKPSATSLAYLYTS 1720
Db 1596 GPALTV-----TEP---VAEAGLSYGDA--DLGDELGRPVHGAHP-----ATYITS 1639
QY 1721 GSTGRPKVMIEHVRITVTSGCIPNY-----PSETRMAHMATIAFDGASYEYI 1770
Db 1640 GSTGRPKGW-----VPRGALDNFLADMGRRTPGSGDRLLAVTTVGFDIAGLEIF 1690
QY 1771 SALLFGRTLVCVDYMTTLIDARLKDVFREHVNAASHVT-----SSSQDV 1815
Db 1691 LPLHGAVLVLADEETARDPHALL-----HRVSASGITMVOATPSLMQGVAAVAGDELA 1744
QY 1816 PLRV-----PRRLSRPLMEFFLVVTDSTAPDALDAQGLYQGVQYNGYGTENGVMST 1868
Db 1745 GVRVLGGEALPSELARAL-----TD-----RARSVTNLYGTEATIWT 1784
QY 1869 IYPIDSTESFNGVPIGRALNNSGAYVDPQQLVIGVMGELVVTGDLARGYSXK--AL 1927
Db 1785 --AADVAES---GPVIGRPLANTSAVILDSALRPVPGVPGELYLAGELQALQYHLRALP 1839
QY 1928 DENRFVHITVNDQTVKAVRTGDRVYRIGDGLIEFFGFRMDQFKIRGNRIESAELALL 1987
Db 1840 TSEFRTPADYGPAGTRMTRTGDLV--CRRRDGALRYLSRVDOQVKLRGFRIELGELEALS 1898
QY 1988 RDSVVRDAAVVLQONDOQAPEILGFPVADHDSNKQGSANQVQEGWQDFHESGMYSDIG 2047
Db 1899 RHPAVAESAV----- 1908
QY 2048 EIDPSTIGSDPKGWTSMWDGSOIDPDENHEWLGETTRTLHDNRSLGNVLEIGTSGSMILF 2107
Db 1909 -----TVREDRPG----- 1916
QY 2108 NLDGRLESYVGLERPSRAAFVNKATESIPSLAGKAKVOVGTATDIGQVDDLHPDLVLVN 2167
| | | | |

Db 1917 --DRELGVV----- 1924
QY 2168 SVIOQFSPSEVLAETADTLIHLPNVORIFFGQDVRSOATNEHFLAARAHTLTKGNATKDDV 2227
Db 1925 ----- 1924
QY 2228 ROKMAELEDMEELLEVPAFFTSLKDRFPGLVEHVEIILPKNMEAVNELSAYRYAAVVHVR 2287
Db 1925 -----VPKGE----- 1930
QY 2288 GSLGDELVLPEKDDWIDFQANOLNOKSLGDDLKSSDRAIMAVSKIPEITAFERQVVAS 2347
Db 1931 ----- 1930
QY 2348 LNSNIDEMQLSTIRSSAEGDSSLSVPDIFRIAGEAGFVEVSSARQWSONGALDAVFHHC 2407
Db 1931 -----GPAGTGPVPSAR----- 1943
QY 2408 CSQGTLVNFPDHHLRGSDLLTNRLORLQNRRIAEVRELRSLRLLPSYMPISNIVULD 2467
Db 1944 -----ELRERLGRRLREFMVPTAFVULD 1966
QY 2468 KMLNANGVDORKELSRRAKVVVKOQTAAPLPTFPPISEVEVILCEATEVFGMKVDITDH 2527
Db 1967 ALPLTPNKRKLDRK-----ALPAPEY----- 1986
QY 2528 FENLGGHSLLATKLISRIDQRLKRVITVKDVPDHPFADLASVIRQGLCLOQPVSDGQO 2587
Db 1987 -----DGEVP 1991
QY 2588 DRSAHMAPRTEETALCDEFAKVLGF--QVGITDNFDLGGHSLMATKLAVRIGHRLDTTV 2646
Db 1992 GR---GPRDPREETLCAFAELGVARVIGDGFGLGGHSLLATRLVSRIRALTALGBVL 2047
QY 2647 SVKQVEDHPVLQALDNLVQSKTNEIVGGRMAEYSPFOLLTTEDPEEFEMASIKQ 2706
Db 2048 SVRQFEFTPTAGLSGALD-----RAAGARAP-----LAARPRE 2082
QY 2707 LELOEIIQDIYPTSQMAKAFDHTTARPPFVPEYIDFPPTSEPDAAGLIKAC--ESLVN 2765
Db 2083 R-----VPLSPAQQRLWFLHQFEGPSATYNMPTALRLSGPVDRAALRAIGDLVLR 2133
QY 2766 HLDIFRTVFAESGELYQVWLSC-----LD--LPQVITEDNINTATNEFLIDEFAKEPV 2818
Db 2134 H--ESLRTVFAADDGGSQVQVLPADRAVGRGLVDVDTAGEVABRVGEAARHAFDLTADIFP 2192
QY 2819 RLGHPLIRFTIITKQTKSMRVIMRISHALYDGLSLEHVVRKHLMLYNGR--SLLP-----P 2871
Db 2193 -----LARLFRVSDTEHV--LLLIHHIAGDGSMAPLARDLTAAYAAACAGAAPDWEPLP 2246
QY 2872 HQFSRYMOYTAD--GRESGH-----GFWRDVIQNTPMTILSDDDTVVDGNDATCKALH 2921
Db 2247 VOYADYALWQREVLDGSDPPDSVAARQALAYWKDALAGLPEQL---ELPTDRPRPATAGYT 2303
QY 2922 LSKIVNIQSQVLGRSNIITQAT-----VFNAACALVLSRESDSKDVVFGVIVSGRQL 2975
Db 2304 GDRIAFTVPTALHARLTTELARSTHSLFMVQAAAFATLLTRLGAGEDIPVGTAVAGRND 2363
QY 2976 PVEYODIVGPCNNAVPRVAHIES--DYNOLLHDIDDOYLLSLPHETIGFSDL-----KR 3028
Db 2364 ATE--DLVGFFVNTVLRTDTCGNPTFRELGRVRERDUAAYAHQDVPFERLVEALNPAR 2421
QY 3029 NCTDWPEALTNFSCCITYHNFEYHPESQEQORVEMGVLTKEVNTMEMDEP----- 3078
Db 2422 SLAHHP-----LYQVMITENN-----TAGAGDRTAPASPDTPDVSGQAGALL 2463
QY 3079 -----LYDLATA--GE--VEPDG--AGLKVTIVIAKTQIFGRKRVRHLEEV 3118
Db 2464 NATRMTAGTGVAKFDLALTFRGHRDPSGGPAGMRGSLEYRTLFDRETAEVSIRL 2519
| | | | |

Db 1830 LPEVNVQYKDYVWHTFNQSAAMKQEA-----YWLVDVYRIPSKLDPPDYKRR 1880
 QY 251 -----LSDLHVPNTTAEHRITFPPLSQKALSNAICRTALISLLSR 293
 Db 1881 HDTFEGSSVLEMERELSDHI-----RKLAINGT-----TLYTVMLSAYVILLNK 1927
 QY 294 YTHSDALFCAVTEQSLPFPKHLYAD--GYQTVAPLRVHCQNLRASDVND-----AIS 346
 Db 1928 YTNQTDIVGTGAAGRLHPD---LQDVGVFVNTLALRNEVDTSYSEKFELOQTKERTIA 1984
 QY 347 SYDDRGLHAPFG-----LRDIRTGDNGSAACDFQTVLLVLTGDGSHVN-----GINGFL 396
 Db 1985 AFON---SEYFPDDLRLKLVGRESNRNPL---FDTMEVLEDAFMTQKQGVKLSPII 2037
 QY 397 QOITESHENPCN-----NRALLHCOMESSGALLVAYYDHNVDTSLOTTLLLOQFGLHIK 452
 Db 2038 FELDNAKFIMFNWDFEQKIVLNIESTS-----LFKDETI-----QKIAEDYFRILE 2086
 QY 453 CLOSPLDLSMAEVLNMTYED-RAETESWN-SOPLEVQDTLIHHEMLKAVSHSPTKTAIQ 510
 Db 2087 EVSENLDV-ALHODIMISROEKHTLLESFNHTKATYKPAKHOLFEQAKRIPDHTAVV 2145
 QY 511 AWGDWYSELNVNSSLAVHISGLGRAQAAIIPVYFEKSKWVIASMLAVLKSGNAFTL 570
 Db 2146 FEDOKLYRQOLNEKANQVALLREKGVK-PDVLVGIMERSSDMIAAILGVLKAGGAYLP 2204
 QY 571 IDPNDPARTAQVVTQTR-----ATVALTSKLHRETVQKLVGRVVDDELLOSVASDD 625
 Db 2205 IDEYPERMRYNAFSEVKVVIISDVPLABELTAESIE-----LIHMDDERIAGODRS-D 2258
 QY 626 FSSLTKSODLAVYFTSGTDPKGMIEHRAFPSSCALKFGASLGINSIDRALQFCGHAF 685
 Db 2259 IDNVNOSGDLAVIYVTSSTGTRPKGVNIEHQSLINLCSWHQSCFVQONDNSIYASISF 2318
 QY 686 GACLLBIMTTLNGCVCIPSDDDRMNSIPSFNRYNVNMMATPSYMG-----FSPED 740
 Db 2319 DAFVWELFPYITAGATVHVLNQETRLD--VEKLNRYFHDHII-TISFLTPVCEQFTALD 2375
 QY 741 VPLGATLVLGVEGSSNAIWAAPKLQALLNGVQSESSIC--FASNMTSTEPNMGRVAG 798
 Db 2376 NISLRTLTLGGDKL-----NVFEKSYIVNNGTPTNVVATSFPIDKSHQNIPIKPID 2431
 QY 799 AHSWIDPNDINRLVPIGAVGELVIESPGIARDYIVPPPEKSPFPTDIPSWYPANTFPD 858
 Db 2432 NVKVIYLNKDL-QCLPLGASGELCIAGELGARGVYVNRPELTREKFI-----GNPVP 2482
 QY 859 GAKLYRTGDLARYASGSIVCIGRIDSOVKIRQORVELGAIETH-LRQOMPDDLTITVREA 917
 Db 2483 GERMVYRTGLAKMLPDGNIQFGRVDOQVKIRYRIEPEGEIENRLLKYEKIEBAVIA-- 2540
 QY 918 TKRSQSANSTSLTAFILIGSSYFGNRPSDAHLHDHDKATKAINIKLEQVLPRHSTPSYICM 977
 Db 2541 --REDGDHPYLCAVY-----VKEVEPEKIRAFLLKSLPDYIMQIFVOL 2585
 QY 978 LELPRTATGIDRRRLRIMGKDI-LDKQTOGAIVQOAPAPIVFADTAALKHSIWQSIG 1036
 Db 2586 DGLPLTVNGKVBKSLPVPERSVTMDR-----YEAPRD-----QMEKLVSIWEEALG 2634
 QY 1037 IDPATVNVGATFELGCGNSITAKWVN-MARSVGMDLKVSNIYQHTPLAGISAV-VKGGP 1094
 Db 2635 IN-KIGINSHPFEAGGHSKAAALYSTHKELVNKLPLRQIFETPTIKGLRDISVRRRK 2692
 QY 1095 LSVTLIPKSTHEGPVEO-SYSOGRLEWLDOLDVGS-LWYLIPYAVMRGPVNVDALRRAL 1152
 Db 2693 CFY--IDRTEKPYRILSSAKRRIILSQ--TGSHVYMPFAMTLEGDFDIRFENTL 2748
 QY 1153 AALQORHETLRTTFEDQDGVGVQIVHEKLSSEMKVIDLCSGLDPPEVLNQOQTTPFNLS 1212
 Db 2749 KMWKRHESFRTSFVIMDEVNQOIKEIDFQVAYSDI--CKESAEEKIKSFIRPFHLE 2805
 QY 1213 SEAGHRATLLRLGEDDHILTIYVHHIISDGWSIDVLRDLNOLYSAALKDSKDPISALTP 1272
 Db 2806 KAPLLRAEVVKNLREHLLMFDMHHIISDGVSTDFIQELGALYEG-----KSLKP 2856

QY 1273 LPIQYSDFAKWQDOQFTEQE--KOLNYWKOLKODSSPA-KIPTDFARFALLSGDAGCVHV 1329
 Db 2857 FHIQYKDYAEWENSHARSEELKRQEYWLKTYKGDIPVLDLPIDHKRPLTKSSBGDTVTA 2916
 QY 1330 TIDGELYOSLRAFECNEHNTTSFVLLAFAAFAHYRLTAVEDAVIGTGTPIANRRELEDI 1389
 Db 2917 AIESEFRKLOHMAKENGVTMYMLLAGYTALLSKYTGQEDIIIVGTTPAAGRHHEDIOH 2976
 QY 1390 GCVNTQCMRINTDHDHFTGTLINQKATTTAFAFENEDIPFERVVSALQPCSRDLSSTPL 1449
 Db 2977 GMEVNTLAI RNHEGKKTFRDYLOEVKENTLOAYENQDYFEELEKVN-KRDMARNPL 3035
 QY 1450 AQLIFAVHSOKDLGRKFKQGLSEVPVPSK-AYTRFDMEFHFLFOETDSLKGSYNFADELFX 1508
 Db 3036 FDTMLVYHN-TDVKPFPEAGLRSLRLEIKRIGSKFDITVTASEAADGLRLEVEYSTLFLN 3094
 QY 1509 MEIVENVVRVFFETILRNGLOSSRTPVSILPLTGTIVTLEKLDVL-NVKHYD--YPRESSL 1565
 Db 3095 KERMERLEHILSLE---QAADHPDIAINOIDLTKGERHRLYDFNRTDGVFCREMTI 3151
 QY 1566 ADVFOTQVSAYPDSLAVVDSSCRLTYTELDQSDILAGWLRRRRSMAPAEITLVAFAPRSCE 1625
 Db 3152 PELFEQAEKTPDHPAVAFGDETISYRELNERANSLAFTLRQKGVGPDVITAGILTERSE 3211
 QY 1626 TIVAFPGVLKANLAYLPDVRSPSARVQDILSGLSGPTIVLIGHDTAPPOIEVTNVEFVR 1685
 Db 3212 MIVGIMGLKAGAYLPIDPAYPOERISYIVKD-SDVSVLCAAGDVDPG--EAYTGDIIIR 3268
 QY 1686 IRDALNDSNADGPEVTEHDSKTPSATSLAYVLTSGTSGTRPKGVMTIEHRYI---IRIVTS 1742
 Db 3269 IDOT---GONDHVENLKH-D-IRKQ--HLAIVITSGTSGAPKGMVIEHHVNNLVHGLNE 3322
 QY 1743 GCIPNYPSETRMAHMATIAFDGASYEITYSALLFGRTLVCVDMYMTTIDARALKOVFFREHV 1802
 Db 3323 RIYQHLDAHLNALVALVAPYIFDASVKQIFAALLFGHTLCIVPRETANDAMSLIEYSKNI 3382
 QY 1803 NAASHVTSQDQVPLVRPRRLSRTLMEFF----FLVYTDSTAPDALDAQGLYQ--GVOC- 1854
 Db 3383 N-----VSOMTPAHLNMLAYVDKTELEFVDELIVGVDALTPDVIQ--GLFHFKNLSCN 3435
 QY 1855 -YNGYPTENGVMSTIYPIDS-----TESFINGVPIGRALNNSGAYVDPQOQLVGIVM 1908
 Db 3436 ITNVYPTCECCDAASHQIESGKVPQTPS----IPGRPLNLTISYIVDKELRPLVGIA 3491
 QY 1909 GELVVTGDLARYSDK-ALDENRFVHTVNDQTVKAYRTGDRVRYRIGDGLTEFFGRMD 1967
 Db 3492 GELCIAGEGVARGVYVNRPELTAEKFDVHPF-EPGKKMYKTGDLAMW-LPDGQIEFLGRAD 3549
 QY 1968 TQFKIRGNRIESAEIEAALLROSSVRDAAVVLOQNEQAPETILGFVVADHDHSENDKQS 2027
 Db 3550 HQVKIRGYRIELGEVEEQQLTHEKIEKAAVIAKQDQNGNSYLCAYTASOKELPAADVRF 3609
 QY 2028 ANQVEGWQDHFESEGMYSIDGIDPSTIGSDFKWTSMYDCSQIDFDEMHEWLGCTTTLH 2087
 Db 3610 LER-----EMPDYMPSYF-----VKLDLPR----- 3631
 QY 2088 DNRSGLNVLEIGTGSMLFNLDLSRLSEYVGLSEPSRAAFVKNKATESIPSLAGKAKVQV 2147
 Db 3632 -----TPSGKV-----DRSA-----LPEADGNVNVME 3653
 QY 2148 GTATDIGOVDLHPLDVLVLSVIQYFP-----SSEYLAEIADTLHLHPNVQRIFF 2197
 Db 3654 GTGYD-----PPRNEIERKLVQVREILGAEDIGISHHFFAAGGDSKALQIVSRL-- 3704
 QY 2198 GQVRSQATNEHFLAARAIHTLGNKATKDDVRKMAELEDMEBELL-VEFAFTSLKDRFP 2256
 Db 3705 AKMNLKLEMKALFANPKIKDLSRFITEETPRHKNPKVGTETELLPIQRYFANKEELD 3764
 QY 2257 GLUHEHVEILPKWNEAVNEL-SAYRYAAVHVHVRGSLDELVLVPEKDDWDIDFQANLQKS 2315
 Db 3765 HFNQSFMLFRKQDGYDENIVRTAFNKILEQH-----DALRMTYEEKDGOIIOYNRGYR 3818

Qy	2316	LGOLLKSSDAAMAVSKIPPEITAFERQVVVUNSNIDEMQJSTIRSSAPGSSLSVPDI	2376
Db	3819	LFDL-----DV-----YDRFGFSQ-----EKVPFELATGIOKQSSI-----	3850
Qy	2376	FRIAGEAGFRVEYSSARQWSONGALDAVHHCCSQGRTLVNFPFTHHLRGSDLLTNRPLO	2435
Db	3851	-----RKGLVHUGIFRAEGDHLLIAI--HHLVVDG-----	3880
Qy	2436	RLQNRRIAEIVRERLSLPSYMPISNIVVLDKMLNANGKVDK-KELRRRAKVVPKQK--	2493
Db	3881	--VSWRLIFDEFTL-----YLOALKEPLDIGYKTSYOFARQLKYYAQR	3927
Qy	2494	-----TAAPLPTPPISEVEVILCE-EATEVFGMKVDITDHF-----FNLGGH	2534
Db	3928	LLKEREYWKALEADVPFPAEKLELDTEHSAATLSIRIGCPDVTAKLLRANAFKAYNTEIN	3987
Qy	2535	SLLATKLISRIDQRLKVRITVKDVFDPFADLASVIRLOGLGQOPVSDGOGODRSAHMA	2594
Db	3988	DILLITALIA-----AVRDIICE-----NKLKVMMEHG--REDILGV-----	4023
Qy	2595	PRTEETAILCDEPAKVLGFQVGITDNFFDLG-----GHSLMATKLAVR-----IGH--	2640
Db	4024	-----DITRTIGWFTVVPVFDLGEKEJISQNIKMVKALKRIPKNGIGYGV	4072
Qy	2641	RLDITTVSVKDVDPHPVLFOALIALDNLVQSKT-----NEUVGR-----	2679
Db	4073	KYMTTELQKTQOAPLSFNFGEMDMNRKVPFSQSPFSGESIGGKIVRHCAIEMNAIS	4132
Qy	2680	-----EMAEYSPF-----QL--LFTEDPEFM-----ASEIKPQ-----LE	2708
Db	4133	LNGLITYTTFNODQOTSTIEGLNOSFKLENLEKIVDHCVDKSGSDMTPSDYGDVSLGLE	4192
Qy	2709	LOEIIOD-----IYPTOMOKAFLE---DHTTARPRPVPVYIDFPSTSEPDAA	2755
Db	4193	ELELIKDKYSAFOLEKIYPLANNQKMLFHNAMDQTSG---AYFQIVIKLGRVHFDI--	4248
Qy	2756	LIXACESLVNHLIDIFRTVFA-EASGELYQVVLSCDLDPQIVITENINTATNFEI	2814
Db	4249	LEESFHEIVRRHILRASFEYEITAPRQIIARDRKTPTSIDLTGENRTRQHRFIETVL	4308
Qy	2815	KEPVRIG-----HPLRFTTIKOT-KSMRVIMRISHALYDGLSLEHVVRKLHMLY-----	2863
Db	4309	KEOQKGFDSLSSALMRVCLIKMSDESRYLWISHHHILLDGCWLGIVLSFLFSLYGKI	4368
Qy	2864	-NGRSLPPPHQFSRYMQY-TADGRESGHGFWRDVION-----TPMTILSD	2907
Db	4369	GESRRUKEPKPYDGIYKLEKQDOEEAVAWKYDLKYESRSELPFAFNRCATSEYCY	4428
Qy	2908	TVVD-GNDATCKALHLSKVINPISQVLRGSSNIITQATVFNAAALVLSRESOSKDV	2966
Db	4429	KVIFSFKELTKITRIAK-----QHHVTINTVLQGTWGMILAKYKNTDEV	4476
Qy	2967	RIYSGOGLPVE-YQDITVGPCITNAVVPRAHIESS-DYNQLLHDIQOYLLSLPHE	3024
Db	4477	TVVSGREA-PVDGIEEMVGILFIHTIETRISFEGARSFKEVLKKTQAESIESNRYS	4535
Qy	3025	DLKRNCITDWPPEATNFSCCITYTNHFYHPESQEQORVEMGVLTKFVNTEMDPL	3084
Db	4536	EIQVLSMKRELITH--VMAFQNYAF-DEELFRSQSGEGTFFLEGVH-GKERTN	4590
Qy	3085	AGEVEDPDGAGLV-----TVIAKTQLFGRRKRVHEHLLLEEVSKTFEGUN	3126
Db	4591	TGVLEDEQLKLKLTFFNENVYDNIITETLEKHIIITVAEOVAEDETOTLR	4641

A:Title: Characterization of the syringomycin synthetase gene cluster. A link between
A:Reference number: Z18153; MUID:99047670; PMID:9830033
A:Accession: T14593
A:Status: preliminary; translated from GB/EMBL/DDBT
A:Molecule type: DNA
A:Residues: 1-9376 <GUE>
A:Cross-references: EMBL:AF047828; NID:g3510628; PID:g3510629; PID:AAAC80285.1
C:Genetics:

Query Match	11.6%;	Score 1876.5;	DB 2;	Length 9376;
Best Local Similarity	21.9%;	Pred. No. 8.8e-100;		
Matches	813;	Conservative 523;	Mismatches 1202;	Indels 1171; Gaps 120;
Qy	267	RITFPLSQKALNSAICRTALSIILLSRYTHSDEALFGAVTE-----QS	309	
Db	3434	RIRTOAERGMSPAVLFHVAWAQVMARCTGRDDVVTGTRVLTGTAGTAGAERAMGPMNT	3493	
Qy	310	LPFDKHYLADG-----TYQTVAPLRVHQSNL-----RASDVMDAISYYDDRLGHUAPF	358	
Db	3494	LPVRVQLTQGAQELVMATHRDLSELLSHSEQASLALAQRCSSVATGVPFESSLLNY----	3549	
Qy	359	GLRDIRVTGNGSNAACDFQIVLLVTDGSHVNN-----GINGFFLOOITESSHEW-PCNNRA	412	
Db	3550	-----RHOGEDNRLQWPGMRL-----DGTERTNYPCLUSVNDYSELDLIITHSQPANPQR	3601	
Qy	413	L--LLHQMBESSGALLVAYVDHNVIDSLQTRLLQOQFGLIKQLQSPDLSSMAEVLNMT	470	
Db	3602	LCAMMQCALE-----QLTDLAHTPQMAVTQDVLPAERNLL-3639		
Qy	471	EYDRAETESNNSOPLEV-QDTLIHHEMLKAVSHSPKTKATAOANDGDMVTSELDONVSRLA	529	
Db	3640	-----LETENQTRQDYPTDLCIQHLEFAQVTRTPDAIAVAVOQRLSYSLADLNRAQNRLA	3693	
Qy	530	VHIKSLGLRAQOALIPVYFEKSKWVIASMLAVLKSGNAFTLIDPNDPARTAQOVTQTRA	589	
Db	3694	HLHLSLGIQVDDR-VAICVERGEVEMTGLGLVGLKAGAAYVPLOPAERLARLAWITDSQP	3752	
Qy	590	TVALTSLKHRETQVKLGRVCVVYDDELLQSVASDDFSSLTK--SQDLAKVIFTSGTSGD	647	
Db	3753	AALLTFLGLQDRLPALSMPLVLLDDEQYQGLAECDONPVVPTTLGVRNLAVVITSGTGN	3812	
Qy	648	PKGIMIEHRAFFSCALKFEGASLGINSOTRALQGTTHAFGACLEIMTTLINGCVC-----	703	
Db	3813	PKGVMIEHRLGVNSVDAARLFDLSQSDTVLQQNTLNFDSLSEVEIFPALLAGATLAPSRE	3872	
Qy	704	-IPSDDDRMNSI-PSFINRYNNVN--NMATPSYMGTFSPEDVPCLATLVLVGQMSVVN	759	
Db	3873	IFGSEGTETHGIQPTVLHLHTAHWHHTLVAEWHNQPOAAEORLQHVRLINVTGDALSAQKL	3932	
Qy	760	AIW-----APKLQLLNGYQGSSESSICFASNASTE-----PNMNGRA-----VGAHSWVID	805	
Db	3933	KLWDEVREPAHTLLINTGPTTEATVSCYAAVYSYVDAAGSEGSNGATIGKPMANTRYLLD	3992	

Db	6083	ATSWEIFDVL	PD	TL	SL	PI	GP	PN	AN	VR	VY	LD	AR	RE	VP	MG	VG	VD	IY	IG	GT	GA	LV	LN	RR	6142																							
Qy	2363	--SAE--	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	2394																							
Db	6143	ELTAERSE	DP	F	SQ	AG	AR	LY	RT	GD	LA	WL	PD	GN	LE	YA	RD	GN	GQ	V	K	R	G	F	VE	---	6196																						
Qy	2395	SQNCALDA	VEH	-----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	2441																							
Db	6197	--GEISV	I	HL	VD	GN	SV	VV	VA	HE	AS	PG	DR	IL	V	Y	T	V	H	-----	----	----	----	----	----	6241																							
Qy	2442	IAIEVRER	LR	SL	LP	SY	MP	IS	NI	V	LD	K	MP	NA	NG	K	VD	R	K	EL	S	RR	AK	V	V	PK	Q	TA	AP	LP	T	F	2501																
Db	6242	--EDLRA	Q	L	SAD	LA	EY	MP	S	V	F	VR	LD	AL	P	L	T	LN	G	K	V	D	K	A	L	-----	6280																						
Qy	2502	PISEVE	IL	CEE	ATE	VE	FC	MK	V	D	T	D	H	F	N	G	G	H	S	L	AT	K	L	I	S	R	I	D	Q	R	K	V	R	I	T	V	K	OV	F	D	H	2561							
Db	6281	-----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	6280																							
Qy	2562	PVFADLA	S	V	I	R	G	L	Q	P	V	S	D	G	O	G	O	R	S	A	H	M	A	P	T	E	T	E	A	I	L	C	D	E	F	A	K	V	L	G	F	O	-----	2620					
Db	6281	-----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	6322																							
Qy	2621	FFDLGGH	S	M	T	K	L	A	V	R	H	L	D	T	T	V	S	K	V	D	F	H	P	V	L	F	O	L	A	I	A	L	D	N	L	V	S	K	T	E	I	N	E	I	V	G	R	E	2680
Db	6323	FFELGGH	S	L	A	A	Q	L	V	S	R	Q	L	G	M	A	L	R	O	L	F	N	H	P	T	V	A	E	L	A	K	V	D	G	L	T	V	D	S	D	I	-----	EP	6379					
Qy	2681	MAEYSP	Q	L	L	T	E	-----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	2731																							
Db	6380	IERNAP	L	A	S	F	S	O	O	R	L	F	D	R	D	F	G	A	S	S	A	H	M	P	S	L	L	-----	6418																				
Qy	2732	TARPRP	P	V	F	V	I	D	P	S	T	S	E	P	D	A	G	L	K	A	C	E	S	V	N	H	I	D	I	F	T	V	F	A	E	A	S	G	E	L	Y	Q	V	L	S	C	L	D	2791
Db	6419	-----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	6450																							
Qy	2792	PIQVIE	-----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	2834																								
Db	6451	PQVIAA	A	D	S	G	F	A	L	E	D	L	S	S	Q	P	Y	E	Q	A	S	N	A	S	I	A	D	S	E	A	A	P	F	D	R	G	P	L	I	R	G	L	L	R	A	D	6510		
Qy	2835	SMRVIM	R	I	S	-----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	2889																								
Db	6511	DEHML	L	T	O	H	I	I	S	D	G	S	V	G	L	I	N	E	F	T	A	L	Y	Q	A	E	T	Q	R	P	D	P	L	P	A	L	S	I	Q	Y	A	D	A	A	Q	R	T	F	6570
Qy	2890	G-----	----	----																																													

RESULT 12

probable non-ribosomal peptide synthetase Pr2402 [imported] - *Pseudomonas aeruginosa* (strain H83345)
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: F83345
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Brinkman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.

Db 2123 PLPEVAGETLAYVIYTSSTGQPKGVAVSQAALVAHQAAARTYGVGPGDQLOQFASISF 2182
Qy 686 GACILEIMTWTLINGCVCIPSDDDRMNS--IPSFINRYNNVMNMATSYMGTSPSE---- 739
Db 2183 DAAAEQLFVPLLAGARVLL--GDAGWSAQHLADEVERHAVITILDPLPAYLQQAEELRHA 2241
Qy 740 -DYPGLATLVLVGEOMSSSSNAIWAPKLQL-LNGYQGSSESSICFASNMTPEPNN--MG 794
Db 2242 GRTIARACILGGEAWDASLLTQOAVQAEAWFNAYGTEAVITPPLAWHCRTGEGGAPAG 2301
Qy 795 RAVGASHWIDPNDINRLVPIGAVGELVIESPGIARDYIVPEPPEKSPFFTDIPSWYPAN 854
Db 2302 RALGARRACILDAALQPCAP-GMIGELYIGGQCLARGYLGRPGGTARFVADPFS----- 2355
Qy 855 TFPDGAKLYRTGDLARVADSGVLCGRIDSQVKIRQORVELGAGIETHLRQOMPDDITIV 914
Db 2356 --GSGERLYRTGDLARYRVDQVEYLKQADQKIRGFRIEIGIESQLLAH-----PYV 2408
Qy 915 VEATKRS-QSANSTSLIAFLIGSSYFCNRPDSAHILDHDATKAINIKLEQVLPRHSIPSF 973
Db 2409 AEAVALDGVGGLLAAYLVG-----RDA-MRGEDLLAELRTWLGLRLPAYMQPTA 2459
Qy 974 YICMLELPRATGKIDRRRLRIMGKIDLKQTOCAIVQAPAPIPVFPADTAAKLHSTWQ 1033
Db 2460 WQVLSLPLNANGKLDKAL-----PKVDAARRQAGEPPREGLEERS--VAALWEA 2508
Qy 1034 SLGIDPATVNVGATFFELGNSGTAIKVNNAR-SVGMOLKVSNIYOHPTLAGISAVVK 1092
Db 2509 LLGVE--GIARDEHFFELGHSLSATRVVSRRLQDLPLRLLFRPVLADFAASLES 2566
Qy 1093 DPLS---YTLIPKSTHEGPEVSQYSGRLWFLDOLDVGSWLVIPIYAVRMGRPVNDAL 1148
Db 2567 QAASAPVLQILPR-VAELPL--SHAQORMWELWKLEPESAAVHLPSVLHVRGVLDQAL 2623
Qy 1149 RRLAALAEORHETLRTTFEDQGVGVQIVHEKLESEMKVYDLQCG-SPLDPFEVLNOBOTT 1207
Db 2624 QOAFDWLVLRHETLRTFEEVDGQARTILANPLRIVLEDCAGASEATLRQRAVEIRQ 2683
Qy 1208 PFWLSSEAGWRATLLRLGEDDHILTIVMHIIISDGSIDVLRDLNOLYSAAKDKSDPL 1267
Db 2684 PFDLARGPLRLVLLALAGOEHLVITQHIIISDGSVMQVMDLLOQVAAAARREGQ--- 2740
Qy 1268 SALTPLEIOYSDFAKWKQOFIOE--KOLNYWKKOLKDDSSPA-KIPTDFARALLSGDA 1324
Db 2741 PTLAPLTQYADYAAHRAWLDSGEARQDYWRERLGABOPVLEPADRVPAQASGRG 2800
Qy 1325 GCYHVVTIDGLOSLRAFPCNEHTTSFVLLAAAFRAAHYRLTAVEDAVIGTPIANRNPE 1384
Db 2801 QRLDMALPVSLSEALLACARREGVTPFEMILLASFQVLLKRYSGSDIRVGPVIANRNAE 2860
Qy 1385 LEDIGCFVNTQCMRINIDHDTFGTLINOVKATTTAAFNEDIPERVVSALQPGSRDL 1444
Db 2861 VERLIGFFVNTQVLRQCVDAGLAFRLDGLGRVREAAALGAHQDLPEQLVDALQP-ERNL 2919
Qy 1445 SSTPLAQLFVAVHSQDKLGRFKQGLSEVPVPSKAY----TREDMEHFLFOETDSLKGSV 1500
Db 2920 SHSPLFOVMNHQS-----GERQDAQVDGLHIESFADWGAAAOQFDLALDTWETPDGLGAAL 2975
Qy 1501 NFADELFRKMETBVNVVYFTEILRNGLOSRTSPVSLPLTDGIVTLTKLDVLNVKHWDPY 1560
Db 2976 TYATDLFEARTVERMARHQNLLRGMLENPQASVOSLPLMDAERQQLLEGNNATAEYP 3035
Qy 1561 RESSLADVQTOVSAYPDSLVAVDSCRITYTDLDRQSDILAGLWRRRSMAPTLLVAVFA 1620
Db 3036 LQGVHRLFEQVETPTAPALAFGEERLDYAEELNRRANRAHALHIERGVGVADRLVGAM 3095
Qy 1621 PRSCETIVAFGVLKANLAYPLDVRSPSARVQDILSGLSGPTIVLIGHDTAPPDIEVTN 1680
Db 3096 ERSIEMVVALMAILKAGGAYVPDPEYPERQAYMLEDVSGVQLLLSQSHKLPLAOGVOR 3155
Qy 1681 VEFVIRDALNSDAGFVEVIEDHSTKPSATSLAYVLYTSGTGPRKPGVMIHRIIRTV 1740
Db 3156 IDLDQ-ADAWLENHAENNPICIE-----LNGENLAYVIYTSGSTGKPKGAGNRHSAL- --- 3205

Qy 1741 TSGCIPNYPSETRMAHM-----ATIAFDGASYEYISALLFGRTLVCVDYM 1785
Db 3206 -----SNRLCWMQOAYGLGVGTVLQKTPFSDVSWVEFWFLMSGARLVAAAPG 3255
Qy 1786 TTLDARALKOVFFREHVNAAASHVTS-----SSODVPLRVPRLRSRTLMFFFLVVVTDSTA 1839
Db 3256 DHRDPAKVALINREGVDTLHFVPSMLQAFLOQDEDV-----VSCSTSLKRVCSGEALS 3308
Qy 1840 PDA-----LDAQGLYQGVOCYNGYGTENGVMSTIYPIIDSTESFINGVPIGRALNNSG 1892
Db 3309 ADAQQQVFAKLPOAGL-----YNYLGTAAIDVTHW--SCVEEGKDAVPIGRPIANLG 3360
Qy 1893 AYVDPPEQQLVGIVGMGELVVTGDLARGYSDK--ALDENRFV--HITVNTQTVKAYRTGD 1949
Db 3361 CYILGDGLPEVPVGVGLGELYLAGRLARGYHQRPGLTAERFVSPFVAGE--RMYRTGD 3417
Qy 1950 RVRYRIGDGLIEFFGRMDTKFIRNGRIESAEIETAALLROSSVRDAAVVIQOONEDQAPEI 2009
Db 3418 LARYR-ADGVIEYAGRIDHVKRLGRLEIETEARLLEHPWVREAAVLAVDSR-----QL 3472
Qy 2010 LGFVVAHDHSDNDKGOSANQVEGWODHFESGMV-----SDIGEIDPSTIGSDFKGM 2061
Db 3473 VGVVLESECGDWRREALAHLATSLPEYVWPQAWLALERMLPSNGKLDKRALPRPQAAA 3532
Qy 2062 TSMYDGSQIDFDEMHEWLGETTRTLHDNRSLGNVLEIGTSGMILFNLDSRLSESYVGLPE 2121
Db 3533 GQTHVAPQNDMERRIAAAMVADVLKLEEVGATDNFFALG-GDSIVSIQVWSKRA-AGIQ- 3589
Qy 2122 SRSAAFVYNKATESIPSLACKAKV-QVGTATDIGOVDLHDPDLVLSVYQFPSSSYLA 2180
Db 3590 -----FTPKDLFOQQTVOGLARVARVAAVQMEQ-----G 3619
Qy 2181 EIADTLILHPNVQRIFFGDVRSQATNEHFLAARAIHTLGKNATKDDVRQKMAELEDMEEE 2240
Db 3620 PVSSEITVLLP-FQRLFF-----EQIPNQHWNQS 3648
Qy 2241 LLVEP---APFTSLDKRFPGLVEHVELPKNMEAVNELSAIRYAAVHVHVGSLGDELVL 2297
Db 3649 LLLKPREALNAKALEAALQALVEHDLALRLRF--HETDGTWHAEE--HAETLGLGALLWR 3703
Qy 2298 VEKDDMIDFOANOL---NOKSL---GDLKS-----SDAAIMVASKIPEITAFERQV 2344
Db 3704 AEA---VDROLSEICEESQSRSLDLADGPLRLSLVDMDAGG-----QRL 3746
Qy 2345 VASLNSNID--EQWL-----STIRSSAEGDSSLSVPDIPRIAGEAGFREVSVSARQWS 2396
Db 3747 LVITHLVVDGVSVRILLEDLQRAYQOSLGEAP-----RLPG-----KTSFFKAWA- 3792
Qy 2397 NGALDAVFHCCSQSGRTLVNFPDTHHLRGS-----DLTNRPL-----QRLQ 2439
Db 3793 -----GR-----VSEHARGESMKAQLQFWRELLEGAPAEPCBHPQGALE 3832
Qy 2440 RRLAIEVREEL-RSLLPSPYMPIPSNIVLDKMPNLNANGVKDRKLSRAKVVVQKQOATAAPL 2498
Db 3833 QREATSVQSRFDRSLTER-----LLKQAPAAIYRTQVNDLLLTALARYVRCWMSGAS-- 3882
Qy 2499 PTFPISEVEVILCEEATEVFGMKVDITDHFNLGHSLLATKLISR-----IDQRLKVRIT 2554
Db 3883 -----SSLVQLEGH-----GREELFADIDLRTVGHFTSLFPVRLS 3918
Qy 2555 VKDVFHPVADLASVIRQGLGQQPVSD--GQGDRAHSAHMAPRTETAILCLDEFAKVLGF 2613
Db 3919 -----PV-ADLGESLKAKEQLRAIPDKGLGYLLRYLAG-----EESARVLG 3961
Qy 2614 --QVGTIDNFFDLGHSMLATKLV-----RIGHRLDTVSVKVDVFDHPVLFLQALD 2665
Db 3962 LFOARTFNY--LQGFDAQFDEMALLDPAGESAGAEADP-----GAPLD 4003
Qy 2666 NLVQSTNETIVGREGMAEYSPFOLLETPDEPEEFMASEIKPOL----- 2707
Db 4004 NML-SLNGRVDFDELSDMSFSSQMEGEOVRRLADDDYVAELTALVDFFCDSPRHGATPS 4062

Qy	2708	-----ELQEIITODIYPSTOMQOKAFLEPDHHTTARPRFPVFPYIDFPSTSE	2750
Db	4063	DFPLAGLDQARLDALPALEE-VEDYIPLSPMOQGMLF-HSL-----YEQASSDYINOMR	4115
Qy	2751	PDAAGL-----IKACESLVNHLIDFTVFAEASGELYQVVLSCLDLPQIVIE-----T	2798
Db	4116	VDVSGLDLPFRFAWQSALDRHAILRSGA-WOGELOQ-----PLQIVYRQROLPEFA	4166
Qy	2799	EDNINTATNE---LHDEFAPKEPVR-----LGHPLIRFTIITKOTK-SMRVIMRISHALYDGL	2850
Db	4167	EEDLSQANRDAALLALAAAEERGERGELORAPLURLLLVKTABGEHLIYTHHHILLDGW	4226
Qy	2851	SLEHVYRKHLMYNGSLPPLPHQF--SRYMQYTA----DGRESHGCFWRDVONTPKTIL	2904
Db	4227	SNAQLSEVLESTAGRS-----PEOLRGYSDYIAWLORQDAATAEAFWRE-----QMAAL	4278
Qy	2905	SDDT-VVDG-----NDATCKALHLSKIYNIPSOVLR--GSSNIITQATVFNAAICALVLS	2955
Db	4279	DEPTRLVEALAPGLTSANGVGBHLREVDATATARLURDFARRHQVLTNTLVQAGWALLIQ	4338
Qy	2956	RESDSKDVFGRTVSGROG-LP-VFYODIVGPTCNAPVRAHT-ESSDYNOLLHDITQDY	3012
Db	4339	RYTGQHTVFGATVSGRPADLPGENQ--VGLFINTLPVVVTIAPQMTLDELQGLQRQN	4396
Qy	3013	LLSLPHETIGFSDLKNCNTDWPEAINTFSCCIYHN---FEYHPESQFQORVEMGVLTIK	3069
Db	4397	LALREQEHPPLFELQ-----W-----AGFGEAVFDNLLVFENYVPDEVLEERSAGGV--R	4446
Qy	3070	FVNIEMDEPL-YDLATA	3085
Db	4447	FGAVAMHEQNTYPLALA	4463

RESULT 13

T31075
 tyrocidine synthetase 2 - Brevibacillus brevis
 C:Species: Brevibacillus brevis
 C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
 C:Accession: T31075
 R:Mootz, H.D.; Marahiel, M.A.
 J. Bacteriol. 179, 6843-6850, 1997
 A:Title: The tyrocidine biosynthesis operon of Bacillus brevis: Complete nucleotide sequence
 A:Reference number: Z20969; MUID:98012987; PMID:9352938
 A:Accession: T31075
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3587 <MOO>
 A:Cross-references: EMBL:AF004835; NID:g2623770; PID:g2623772; PIDN:AAC45929.1
 C:Genetics:
 A:Gene: tycB
 C:Function:
 A:Pathway: tyrocidine biosynthesis
 C:Superfamily: surfactin synthetase; acetate-CoA ligase homology; acyl carrier protein H
 C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
 F:513-957/Domain: acetate-CoA ligase homology <ACL1>
 F:975-1043/Domain: acyl carrier protein homology <ACP1>
 F:1553-1992/Domain: acetate-CoA ligase homology <ACL2>
 F:2010-2078/Domain: acyl carrier protein homology <ACP2>
 F:2990-3025/Domain: acetate-CoA ligase homology <ACL3>
 F:3043-3110/Domain: acyl carrier protein homology <ACP3>
 F:1007, 2042, 3075/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

	Query Match	11.3%; Score 1816; DB 2;	Length 3587;
	Best Local Similarity	22.0%; Pred. No. 5.3e-97;	
	Matches 844;	Conservative 577;	Mismatches 1359; Indels 1048; Gaps 140;
QY	31 SYEQPLHLVGLSSRIEAIKCTPQLMDICNALDKO--SAIGHAVVDVPTDIDISRFA	88	
DQ	: :	: :	: :
DQ	5 SKEQVDMVAL-----TPMGEGLFHALLDQHNSHLVQMVSISLQGDLGVLFT	53	
QY	89 LAWKEIVNQTALRAFAFTSDSGKTSQVILKDSEVF-FSWMCWSSSSDPDEVVRDEAAAAA	147	
DQ	: : : : : : : : : : : : : : : : : : :		
DQ	54 DSHLVIVRDYDFRTLFYEKLKPLOVVLLKORPIPIEFYDLKSACDSESKOLRYTOVKRA	113	

1140 RGPNVDALRRALAEORHETLRTTFEDODGVGVQIVHEK-----LSEMKVIDLC 1191
1094 EGTLEKARLEHATSINVAHESURTSFHTINGEFSVRIRHEQGLHPVIVYETAEEQVNEVI 1153
1192 GSDLPPEVLNQBTTPFNLSSEAGNRATLLRGEODHILTIWMHHIISDGWSIDVLRD 1251
1154 LGFMQPFDLV---TAPL-----CRVLGVLKLAENRHVLIIDMHHIISGVSSQILND 1202
1252 LNOLYSAAKDKDPLSALTPLPIQYSDFAKWOK--DQETEQEKQLNWKQKOLKSDSPA- 1308
1203 FSRLYQN-----KALPEQRHXYKFAVWEKAWTQTDYQOEKYWLDRFAGEIPVL 1253
1309 KIPTDFARALLSGDAGCVHVTIDGELYOSLAFACNEHNTTSFVLLAFAAHRYLTA 1368
1254 NLPMDYPRPAVQFEGERYLFRTEKQLLESLODVAQKTGTLLYMWLLAAHVHLLYSYGQ 1313
1369 EDIVIGTPTANRNRPELIDIGCFVNTQCMRINIDHDTFGTLINOVKATTTAAFNEDI 1428
1314 DVMIGTVTAGRVHPDTEMTGNEVNTIAMRQSAPTFRQFLLEVKDNTLAAFEHQY 1373
1429 PFERVVSALQPSRDLSTPLAQLIFAVHSQKDLGRFKQGLSVF--VPSKAYTRFDMF 1487
1374 PFEELVEKLA1-QRNRSRNPLFDTLFLQNM-DADLIELDGLTVTPVYVEGEVAKFDLSL 1431
1488 HLPQETDSLKGSYNFADELFKMETVENVVRVFEILRNGLOSSRTPVSLPLTDGIVTLE 1547
1432 EASENOAGLSFCFECTKLFARETIERMSLHYLIQILO--AVSANTQELAQI-EMLTAHE 1488
1548 KLDVL---NVKHVDYPPRESSLADVFQTSAYPDSLAVVDSSCRLYTELDROSIDLAW 1604
1489 KOELLVHFNDAALYPAESTLSOLFEDQAQKPEQTAVVFGDKRLIYRELNERANQLAHT 1548
1605 LRRSPMAETLVAVFAPRSCETIVAFVGVUKANLAYPLDVRSPSARVODIISGLSGPTI 1664
1549 LRAKGVAQSGVIMAQRSLEMAIGITAILKAGGAYVPIDPDYPNERIAYMLDCR--RL 1606
1665 VLIHGHTAPDIEVTNVEFVRIRDALNDSNADGFEVIEHDSKP--SATSLAVLVTSGS 1722
1607 VLFOOQLA--EKTANVECLYL-----DEEGSYSPQENIEPIHTAADLAVIITYSGT 1657
1723 TGRPKGMIEHRIIVRTV-----SCIPNYPSETRMAMHATIAFDGASYEYI 1770
1658 TGRPKGMVMEHRIIVNSVNTWRNDEFALSVRDSGTL-----SLSFADAFALTFF 1706
1771 SALLFQRTLVCDYMTTLDRALKVDFREHVNASHVTSQSDVPLRVPRLSRLTMFF 1830
1707 TLIVSGSTVYMLPDHEAKDPIALRNLIAAWECSYVVFVPS-----MF 1748
1831 FLVVTSTAPDALDAOGLXOG-----VQCYNGYGPTEGYMSTIYPI 1872
1749 QALCESTPADIRSIQAVMLGGEKLSPLVOLCKAMHPQNSVMNATGPTESSVMAT-YLR 1807
1873 DSTESFINGVPIGRALNNSGAYVVDPEQOLVGIGVMGELVVTGDGLARGYSK-ALDENR 1931
1808 DTQPD--QPITIGRPIANTAIYVDQHOLLPLPVGVGEICIGHGLARGYWKPKELTAEK 1865
1932 FVHITVNDQT--VKAYRTGDRVRYRTCDGLIEFFGRMDQFKIRGNRIESAETEAALLR 1989
1866 FV---ANPAVPGERMVKTGDLGRW-LHDGTIDFGRVDDQIKVRYGRIEVEGEIEAVLLAY 1921
1990 SSVRDAAVVLOQNE----- 2003
1922 DOTNEAIVAYQDDRGDSYLAAYVTGKTAIESELRAHLLRELPAWVWPTYLIQLDAFPL 1981
2004 -----DOAPEILGFVVADH----- 2017
1982 TPNKGVDRKALPKPECKPATGAAYVAPATEVEAKLVAIWENALGISGVGLDHFPELGGH 2041
2018 -----DISENDKGS----- 2027
2042 SLKAMTVAAQVHREFQIDLLLKOFFAAPTIRDLARLIEHSEQAAGAAIQAPAEQAYPVPS 2101
2028 -----ANQVBGWQDHES-GMYSIDIGEIDPSTIGSDFKGTWSDYGSQIDFDEMHEW 2078

2102 SAQORMYLLHOLEGAGISYNTPGIIMLEGKLDREQLANALQALVDRHDIURTSFEMVGD- 2160
2079 LGETTTLHNRSLGNVLGTEIGTSGMILFNLDLSRLESYVGLPSR---SAAAFVNKATES 2135
2161 --ELVKIHDR-----VAVNME-----YVTAEEQQIDDLFHAFFRPDLS 2198
2136 IPSLAKAKVQVG-----TATDIGQVDDLHPDLVL-----NSVIQY----- 2172
2199 VPPLLRMSLVKLADERHLLYDMHHTAADAASITILFDELAELYQGRELPQEMRIQYKDF 2258
2173 -----FSSSE--YLAETADTLIHL-----PNVQRIFFGCDVRSQATNE----- 2207
2259 VMOKALHESDAFKQOEAYWLSTFAGNITAVDPTDFRPAVKFAGQVTLSDMQDELLSA 2318
2208 -HFLAARAHIT-----LGNKATKDDV-----RQKMAE----- 2233
2319 LHLEAAHTNTLPMVLLAAYNVLLAKYAGQDDLIIVGTPISGRSRAELAPVVGMEVHTLAI 2378
2234 -----LEDMEELLVEPAP-----FTSLKD-----RFPG-----LVE 2260
2379 RNKPATAEKTQKLOEVKQNAL--DAFDHODYPFESLVEKLGIPRDPGRNPLFDTMFIQ 2436
2261 HVEILLPKNM-----EAVNELSAYRYAAVHVHRSGLD----- 2292
2437 NDELHAKTLDQLVRYPYESDSALDVAKFDLSFHLTERETDLFLREYCTKLFKQOVTVERM 2496
2293 -----ELVLPVEK-----DWIDFQANQLNQ-----KSLGDL 2319
2497 AHFLOILRAVTANPENELQEIEMLTAAEKQMLLVAFNDTHREYRADOTIQLFEELAEK 2556
2320 LKSSDAIIMAVSKIPP-----EITAFER-----QVVASLNSNIDEMQLSTKRSSA 2364
2557 MPESHALVFEEKMSFRELNERANQAAVLRERKGVGPAQIVALLVERSAEMVATLATLK 2616
2365 EGDSSLSV-PDI--FRITAGEAGRVEVSSAROWSONGALDAVPHCCSOGRTL----- 2414
2617 AGGAFLPVDPDYPERIR---YMLEDSQAKLVVTHAHL---LHKVSSQSEVVVDVDDPGS 2669
2415 -----VNPPTD-----HHL----- 2423
2670 YATQTDNLPCANTPSDLAYIIVTSGTGKPGVMEHKGVANQAVFAHLGVTPQDRAG 2729
2424 -----RGSDDLNRPLQRLN-----BRIATEVRERLRSLL---PSYMI--- 2459
2730 HFASISFSDASVDMFGLLSGATLYLSRDVINDFORFAEYVRDNATITFLTPTTAYIL 2789
2460 -----PSNIVLDDKMP-----LNANGKVDKRLSRRAKVVVQKQTAAPL 2498
2790 EPROVPSRLTLITAGSASSVALVDKWKKEKTYVYNGYGPTESTVCATILKAKPDE---PV 2845
2499 PTF-----PISEVEVILCEEATEV-----FGMKVDITDHF-----NL 2531
2846 ETITICKPQNTKLYIVDQOLQKAPQMGELCISGLSLARGYWNRPETAELFVNDPVFV 2905
2532 GGHSLAT-----KLISRIDQRLKVR-----ITVKDVF----- 2559
2906 PGTKMYRTGDLARWLDPGTIEYLGRIDHGVKIRGHRVELGEVSVLLRYTVKEAAAIH 2965
2560 --DHPVFADLASVIROGLGL-----OQPVSDGQGDQDSA- 2591
2966 EDDRGQAYLCAYVAVGEATEPAQLRAYMENELPNYVWPAFFIOLEKMLTPNDKIDRKAL 3025
2592 -----HMAPRTEATEILCDFEFAKVLGF-QVGITDNFDFLGGHSLMATKLAURI 2638
3026 PRKPNQENRTEQYAAQPOTELEQLAGIWDVIGIKQVGTQDNFFELGGDSIKAIQVSTRL 3085
2639 GHRLDITVSVKVDVDFHPVLFOLAIALDNLVQSKTNEIVGCREMAEYSPFOLLTEDEEF 2698
3086 -NASGWTLAMKELFOYPTTEEAALRV-----IPNSRE----- 3116
2699 MASEIKPQLEQIIOIIVPSTOMQAKFLFDHTTAR---PRPFVFPYID-FPSTSEPDAA 2754

Db 3117 --SE-----QGVVEGEIATPIQKWHFANNFTDRHHWNQAVMLFREDGFDE----- 3160

QY 2755 GLIK-ACBSLVNHLDFRTVFAEASGELYQVVLSCLDLPQIVETEDNINTATNEF----- 2809

Db 3161 GLVRQAFQOIVEHHDALRWVYKQEDGAIKQINRGLTDERFR-FYSYDLKNHANSEARILE 3219

QY 2810 LDEFKAEVRLGH-PLIRFTTIKQKSHRVIMRISHALYDGLSLHVVVKL-----HMLY 2863

Db 3220 LSDQOSSTIDLEHGLVHFALFKDGDHLLVAIHVGVGVSRILFEDEFSAYSQALH 3279

QY 2864 NGRSLPP-----HGFSRVQVYADGRE--SCHGFWRDVQIQTPTILSDDTFVVDGNAT 2916

Db 3280 QOEIVLPKKTDSFKDRAQLQXYADSDLLRELVAYHNLTTTAAALPTDFTV-----AD 3335

QY 2917 KALH---LSKVINIPSOVLGRSSNIIITQA-----TVFN---AACALVLSRESQKV 2964

Db 3336 RQKHRTLSFALTVPQ-----TENLLRHVHHAYHTEMNDLLLTALGLAVKDWHTNGV 3390

QY 2965 FGRVSGRQGLPVEYQ--DIVGPCTNVAVPRAHIESSDYNQLLHDIQD--QVLLSLPHET 3020

Db 3391 INLECHGREDIQENMNVRTTIGWFTSQYPVLLDMEKAE--DLPYQIKQTKENLRIPKKG 3448

QY 3021 IGFDLKNKCTDW--PEAITNFSCCITVHNPEYHPESQFQORVEMGVLTKEFVNIENDE 3077

Db 3449 IGYELRLTTSQLOPLA-----FTLRPEISFN----- 3477

QY 3078 PLYDLAIAAGEVEPDG--AGLKVTVIATQLFGRKRVHLLREVSKEFE 3123

Db 3478 -----YLGQPSDCKTGFTFSPGLTGOLFSPESERVFLLDISAMIE 3519

RESULT 14

T36248

CDA peptide synthetase I - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 01-Dec-2000

C:Accession: T36248

R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, March 1999

A:Reference number: Z21602

A:Accession: T36248

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-7463 <SAU>

A:Cross-references: EMBL:AL035640; PIDN:CAB38518.1; GSPDB:GN00070; SCOEDB:SCE63.03c

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: cdapsI; SCOEDB:SCE63.03c

C:Superfamily: acetate-CoA ligase homology; acyl carrier protein homology

C:Keywords: carrier protein; phosphopantetheine; phosphoprotein

F:516-1074/Domain: acetate-CoA ligase homology #status atypical <ACLI>

F:1090-1158/Domain: acyl carrier protein homology <ACP1>

F:1715-2184/Domain: acetate-CoA ligase homology <ACLI2>

F:2200-2268/Domain: acyl carrier protein homology <ACP2>

F:2804-3249/Domain: acetate-CoA ligase homology <ACLI3>

F:3265-3332/Domain: acyl carrier protein homology <ACP3>

F:4323-4746/Domain: acetate-CoA ligase homology <ACLI4>

F:4762-4830/Domain: acyl carrier protein homology <ACP4>

F:5363-5786/Domain: acetate-CoA ligase homology <ACLI5>

F:5802-5870/Domain: acyl carrier protein homology <ACP5>

F:6401-6868/Domain: acetate-CoA ligase homology <ACLI6>

F:6884-6951/Domain: acyl carrier protein homology <ACP6>

F:1122,2232,3297,4794,5834,6916/Binding site: phosphopantetheine (Ser) (covalent) #status

Query Match 11.2%; Score 1810; DB 2; Length 7463;

Best Local Similarity 23.3%; Pred. No. 4.6e-96;

Matches 759; Conservative 440; Mismatches 1118; Indels 942; Gaps 107;

QY 46 IEAKPCTPFQ-----LDWIDCNALDKGSAIGHAVYDPTDIDISRFALAWKEIVNQTPA 100

Db 3824 LQDILPLAPLQSGFLFLNYDENARDVY--VGQAFDLGSGFDGTFMRRAAGALLRRHAN 3881

QY 101 LRAFAFTSDSGKTSVILKDSVFWSMCMWSSSSSPDEVVVD-EAAAAASGPRCNRFVLE 159

Db 3882 LRAGFQRTAGTWQVVPAAE-LEPDWRRECDLTDRADEARDAEAGRLAAGDRERRRDLTS 3940

QY 160 D-----MOTKKCOLVMTFTSHALVDVTFQOVRVLSRYFAAYKHEKD--THRPETESS 208

Db 3941 PPLMRFATRLSADRVLVMTNHHILLDGSWPLMQEULTELYVSGGDPVSPVPPVRYRD 4000

QY 209 -----DATDTSQSVSVWSMSEDNAVSAETHWOTHLDNLNASVFPHLSHLKMPN--PT 261

Db 4001 HLAWLGARDRD-----AARDARRSLSGLEAT-----LLAPDAGPA 4037

QY 262 TTAHRITPPLSOKA-----LSNSAICRTALSILLIRYTHSDSEALFGAVTE-- 307

Db 4038 EAAPLGIPEGLDRDATAALSASARGVTMTVVOGAWALAAQATGRDDVVGATVSGR 4097

QY 308 -QSLPFDKHLADGTYQTVAPLRVHCQSNLRASDVMDAISYVDDR-LGHLAPPLGLDIRN 365

Db 4098 PRELFCVESMI--GLFINTLPVRARLDQAEPLGDLFRRLQNEQARLLDHWQV-GLADIOH 4154

QY 366 TGDNG---SAACDFQTVLLVTD--GSHVNNGINGFLQOITETESSHFMPNRRALLHCOMES 421

Db 4155 WAGHGELEFDTAMVFQNPVSADTTSRQLDGLRVAGYDAVESTDF----- 4198

QY 422 SCALLVAYYDHNVIDSLQTRLLQOF-----GHLIKLOSPLDLSSMAEVLNMTYDR- 474

Db 4199 -AVNLVA---HTRDDAL---RLRLDYRADACAGDLVRSI---ADRLRLVLEALVTDSDRP 4248

QY 475 -AEIES-----WNSOPLEVQDTLIHHEMLKAVSHSTKTAIQAWDGDWYSEL 521

Db 4249 VAHDTLDPVRRERVLVEWNGAPTQLPGTPLHELISEQARLTPDAVAVVCDGSLTYAEL 4308

QY 522 DNVSRLAVHRSKGLRAQQAIIPIVYFEKSKWVIAISLAVLXSGNAFTLIDPNDPPARTA 581

Db 4309 DGGANQARHLLEGGLAED-FVAIALAKSLDAVISMLAVLKTGAAYLPIDPDYPAERIT 4367

QY 582 QVVTQTRATVALTSKLHRETQKLVGRCV--VVDDQLLOSVASDDFSSLTQSQDLAYVI 639

Db 4368 YMLDARPALTLTEPY---PVERYTGHVSVTAVTDEERRSPWSA-----RHAAYMI 4414

QY 640 FTSGSTGDPKGMIEHRAFSSCALK---FGASLGINSDTRALQFGTHAFGACLLIMT 695

Db 4415 YTSGSTGRKGVVIEHHALATYLHRAITYTAMTGV-----TVLHSPFLADLTITALTWP 4469

QY 696 LINGCVCTPSDDRRNMSTPSFINRVNVMNMTATPSYMG--TFSPEDVPGLATLVLVGQ 753

Db 4470 LTSGGTVHLTSLEE-SDTQPSLIK-----ATPSHLPLLTLPATSPSHLTLLGGEA 4520

QY 754 MSSSVNAIWA---PKLQLLNGYQSSSSSICFASNSTE-----PNMWGRA-VGAHSWVID 805

Db 4521 LHTDHLVTWRTQHPGQIINAYGPTESTVITDHHVGEDTPDGPVPIGRPFANTQYVILD 4580

QY 806 PNDINRLVPI--GAVGELVIESPGIARDIVTPPPPEKSPFFTDIPSWYPANTPPQCAKLY 863

Db 4581 ----SALRPVAPCVTGELVLAGELARGYLGRPALTAERTAN-----PHSSTP-CARMY 4630

QY 864 RTGDLARYASDGSIVCLGRIDSOVKIRGORVELGAIETHLRQOMPDLTIVVEAT--KRS 921

Db 4631 RTGDLAHNNHHGLTYDGRADHQIKLRGHRIEPEATELTQA-----TGITQATVOLRE 4685

QY 922 QSANSTSLIAFLIGSSYFNGRPSDAHLHDAT-----KAINIKLEOVLPRHSPISPYICM 977

Db 4686 DQGPQRLVAYLV-----VNDSTEYDEKTRDALTSALPDYMWPSALVTL 4730

QY 978 LELPRATGCKIDRRRLRMKGKDLIDKQGTGAIQQAPAPIPVF-ADTAAK-----L 1027

Db 4731 DALPLTPNGKLDORTAL-----PAPAYSASTAGRAPTRPREVL 4768

QY 1028 HSIWQSLGIDPATVNVGATFFELGNSITAIKVMNMR-SVGMDLKVNINIQHPFLAGI 1086

Db 4769 CILFAEVLGDLVLTID--DNFFDLGGHSLATRLVSRTRTALGVLSIQLFETPIVAGL 4826

QY 1087 SAVVKGDPISYTLIPKSTHEGVEQSYSGRULWFLDQLDVGSLWYLIPIYAVRMGRPVNVD 1146

Db 4827 AEALDASGTVTALTARPRPRIPPLSYAQQRLWFLHOLEGSPSYNTVTLRLGLGALDVD 4886
Qy 1147 ALRRALAALEQRHETLRTTFEDQGVGVQIV--HEKLSEEMKVIDLQSGSDLPPEVLNQE 1204
Db 4887 ALRAAISDVVARHESLRTVTFEDERGAYQIVLPVEAASTPTVVDAEEIG--DRLDEA 4944
Qy 1205 QTFPFLNLSRAGWRATLLRGEDDHILITVMHHIISDGSIDVLRRLDNLQYLSAALKDOK 1264
Db 4945 VGHCFDLAQLPARTSLFRVSERHVLVLLIHHIITASDAWSRAPLAQDLTAAYAAVR--SE 5003
Qy 1265 DPLSALTPLTOYSDFAKWKQDQFIEQ-----KOLNWKKOLKD--SSPAKIPTDFAR 1316
Db 5004 APMA--PLTVQYADYALWQOEILGDDTDADSLAGQYALWQKQAGLPEQLDPTDRPR 5061
Qy 1317 PALLSGDAGCVHVTIDGELYQSLRAFNCNEHTTSFVVLAAAFRAAHYRLTAVEDAVIGTP 1376
Db 5062 PAVAGYSGDRVPFTVPTLHRLTELARATNTSAFMVIOQAAVAVLLTRLGAGEDIPIGTP 5121
Qy 1377 IANRRPELEDITGCFVNTQCMRINIDHDTFGLTINOVKATTTAAFNEDIPPERVVA 1436
Db 5122 VAGRTDAAADDLIGLFTNLVLRDTSGDPTFRLLDRVRDLDLAAVAHQDLPPERLVEA 5181
Qy 1437 LQPSRDLSTPLAQLFAVHSOKDLGREF----FOGLE--SVVPVPSKAYTFEDMERHLFQ 1491
Db 5182 LNP-ARTLSHPLFOVLLTNTDHEGALKDISELPGLTVALREVQTSKEDLSFGFAE 5240
Qy 1492 ETDS-----LGSVNFADLKFMTVENVVRVFEILRNGLQSSRTPSVILPTDGIPT 1545
Db 5241 SFTSRRPQGLEAALDFSTELLDRRSQAIAIDRLVLEAVTTAPDRPIGAVELMDPAER 5300
Qy 1546 LEKLDVNLVHVDYPRESSLADYFOTQVSAYPDSLAVVDSSCRITYTELDRQSDILAGWL 1605
Db 5301 ERLVWENGAPQLP-GTPELHISEQARLTPDAVAVVCDGTTLTLYAELDRRANQLARHL 5359
Qy 1606 RRRSMPAETLVAVFAPRSCTEIVAFFGVKANLAYLPDVRPSARVODILSGLSGPTIV 1665
Db 5360 LGEGLRAEDFVAIALAKSLDAVLSMLAVLKTGAAYLPIDIDYPAERTIYWL----- 5410
Qy 1666 LIGHDTAPPDIEVNWFEVRIRDALNSDAGFEVIBHSTKP-SATSLAYLYTSGSTG 1724
Db 5411 ----DDAQPALTLT----APIPPASYSRPTS-EITDVERRSPWSARHAAYMIYTSSTG 5461
Qy 1725 RPKGVMLHVRVIRTVTSGCIPNVPSTRMAHM-ATIAFDGASVEIYISALLFGRGLYCDV 1783
Db 5462 RPKGVJLEHA-LATYLHRRARNTYAMTGVVLHSPFLAFDLTITALTPLTAGTV---- 5516
Qy 1784 YMTLDARALKDVFREHVNAASHVTSQQDVP--RVPRRL--SRTLMEFFLVVTDSTA 1839
Db 5517 HLTSLEAEVQ-----PSLIKATPSHLPLTLTPETASPSHTLIL-----G 5557
Qy 1840 PDALDAQGL-----YGVOCYNGYPTENGWSTIYPIDSTESFINGVPIGRALNNSGA 1893
Db 5558 GEALHTDHLATWRTQHPGAQIINAYGPTESTVNITDHHV-SEDTPOGVPVPIGRPFANTQV 5616
Qy 1894 YVVDPEQLVIGVMGLVVTGDLARGYSK--ALDENREHVLTVDOT---VKAYRTG 1948
Db 5617 YVIDSALRPVAPGTGELYLAGQLARGYLGRFALTAERF---TANPHSSTPGARMYRTG 5673
Qy 1949 DRVYRIGDGLIEFFGRMDTFKIRGNKRIESAEIAALLRSSVRDAVYLOQNEQAPE 2008
Db 5674 DLAWN-HDGLHTYDGRADHOKILGRHRIEPEGETTLTAQTGITQTTVQLRE----- 5725
Qy 2009 ILGFVADHDHSENDKQSANQVQEGWQDFESGMYSDIGEIDPSTIGSKDFKWTSMYDGS 2068
Db 5726 -----DTPG----- 5729
Qy 2069 QIDFDEHWEWLGTTTLHDNRSLGNVLIGTGTGMILFNLDSKRLSEYVGLPEPSRAAAF 2128
Db 5730 -----DQRLVAYL-----V 5738
Qy 2129 VNKATESIPSLAGKAKVQVGTATDIGOVDDLHPLDVLVNLNSVIQYFPSESYLEAIADTLIH 2188
Db 5739 VNDSTE----- 5744

RESULT 15

C69681

peptide synthetase ppsc - Bacillus subtilis

Qy 2189 LPNVQRIFFGDRVSRQATNEHFLAARAIIHTLGNATKDDVRQKMAELEDMEBELLVEPAFF 2248
Db 5745 ----- 5744
Qy 2249 TSLKDRFPGLVHEVVEILLPKNMEAVNELSAYRYAAVVHVRVSGSLGDELVLVPEKDDWIDFQA 2308
Db 5745 ----- 5744
Qy 2309 NQLNQKSLGDLKSSDAAIMAVSKIPEITAFERQVVASLNSNIDEWOLSTIRSAEGDS 2368
Db 5745 -----YDE----- 5747
Qy 2369 SLSVPOIFRIAGEAGFRVEVSSARQWSONGALDAVFHCCSQGRTLVNFPTDHLHLSGDL 2428
Db 5748 -----PT----- 5749
Qy 2429 LTNRLQRLQNRRIAEVRERLRSLPSYMIPTSNIVLDMKPLNANGVKDKKELSSRAKV 2488
Db 5750 -----LRDALASALPDYMRPSAYVTLDALPLTPNGKLDL----- 5783
Qy 2489 VPKQOTAAPLPTPPISEVEVILCEEATEVFGMKVDITDHFENLGHGSHLLATKLISRIDOR 2548
Db 5784 -----TALPAPAYSAS----- 5794
Qy 2549 LKVRITVKDVPDHPVFADLASVIRQGLQOPVSDGQGSARSAHMAPRTETEAICLDEFA 2608
Db 5795 -----TTGTRTPRTPREEILCTLFA 5813
Qy 2609 KVLGFG-VGITDNFFDLGGHSLMATKLAVRIGHRLDITVSVKQVDFDHPVLFQALALDNL 2667
Db 5814 EVLGDLVLTIDDNFFDLGGHSLLATLVRSRARTALGVELSVRQFFETPTIAGLSGAFDR- 5872
Qy 2668 VQSKTNEIVGREGMAEYSPFOLLFTEDPEEPMASEIKPOLELQEIQDIYVSTOMOKAFL 2727
Db 5873 -----AGRARA-----ALTARPRPER-----TPLSTAQORLW 5899
Qy 2728 FDMHTARPRPF--VPFYIDFPSTSEPDAAGLIKACESLVNHLIDITRVFAFASGLYQV 2785
Db 5900 FLHQLEGPSATYNIPTTLRLTGTLDLTD--LQSALNDLLARHESLRTTYTE-DGEGPRQV 5956
Qy 2786 LSCLD---LPIQVETEDN-----INTATNEFLDEFAKEPVRLGHPLIRFTIIKOTKSMR 2837
Db 5957 IHAWEPGLPLGVVDTGEGELDAMLGAGVHHAFDITAGIPVRA-----TLFRISEQEH 6009
Qy 2838 V-IMRISHALYDGLSLEHVVRKHLMYNGRSL--LP-----PHQFSRYMOYTD----- 2883
Db 6010 VLLLLLHHIATDAWSRTPPLGHDLAAYASACAGDVPAWEPLPVQYADYALMQREVLDGEG 6069
Qy 2884 -----GRESHGFWRDVIONTPWTI--LSDD---TVVDGNDATCKALHLSKIVNIPSOVL 2933
Db 6070 DADAPAGROL--AYWTRQLADLPQLDLTPDRPRAVASQDGRVAFSLDADLYVRLTEL 6127
Qy 2934 RGSNNIITQATVNAACALVLSRESKDVVFGVGRVSGROGLPVEYODIVGCPCTNAPVR 2993
Db 6128 ARATHST-FMWQQAALAVLLTRLGAGEDIPIGPVAGRTDDATE--NLGVFFVNTLVLR 6184
Qy 2994 AHIES--DYNQLLHDIDQVLLSLPHETIGFSDL-----KRNCTDWEPAITNFSCCITY 3046
Db 6185 NDTSGNPTFRELEETRRDLAAVAHQDLPPERLVEALNPARTLAHP---LQVVMIL 6240
Qy 3047 HNFYHPESQEQORVEMGVLTFTVNTMEMDEPLYDLAIAGEVDPDGLKVTYIAKTOLF 3106
Db 6241 STAETDPAASLALPGLRVGAERSRLGAAKVDLAFALAEVRDGEGRSTGLTALDFTDLF 6300
Qy 3107 GRKRVHLLLEVKSTFEG 3125
Db 6301 DRSTAKSLVERFVRTLEAV 6319

Query Match 11.2%; Score 1807.5; DB 2; Length 2555;
Best Local Similarity 24.7%; Pred. No. 9e-97;
Matches 681; Conservative 510; Mismatches 1114; Indels 457; Gaps 101;

974	QY	IICMDELPTATGKIGLURKRRKINGKLDLQATGCAIVQQAIPATIPVFPADTAARAKSLWQ	1035
934	Db	MIEMEQWPVPSGKLDNALPAPG--GAADAET-----YTAPRNV-----TEMKLSQWLED	982
1034	QY	SLGIDPATVNVGATFPELGGNSITAIKWN--MARSVGMCLKVSNIOYHPPTLAGISAVVK-	1091
983	Db	VLKNGP--VGIDHNFDFRGGHSLKATALVSRIAKEFQDVQPLKDVFAHPHTVEGLATVIRE	1040
1092	QY	GDPLSYTLIPKSTHEGPEVQSGYSOGRUWFLDQLDVGSLWLIPYAVRMRGPNVNDALURRA	1151
1041	Db	GTDSPYEAIKPAEKQETYPVSSAQKRIYVLOQLEDGGTGYNMPFAVLEBGLKLNLERMDRA	1100
1152	QY	LAALQERHETIRTTFE--DQDGVGVQIVHEKLSSEMKVIDLCGSDLDPEFVLNQEQTT---	1207
1101	Db	EKELIKRHESIRTAFEQDAGGDPVQRIHDEV-----PFTL---QTTVLG	1141
1208	QY	-----PFLUSSBAGWRATLLRLGEDDHILTIYVHHHIIISGWSIDVLRRLDNLQ	1254
1142	Db	ARTEEEAAAAFIKPFDLQSAPLFRQAQTVKVSDEHRLLLVDVHHHIIISGVSNNILIREFGE	1201
1255	QY	LYSAALKDSKDPLSALTPLPIQYSDFAKWKQ-----DQFTEQEKQLNWKKOLKSDSPA-	1308
1202	Db	LYNN-----RKULPALRIQKDYAVWQEGFKTGDAY---KTQGAJWLKQLEGELPVL	1249
1309	QY	KITDPAFPALLSGDAGCVHVTIDGELYQSLRAPCNEHNHTTSFVVL ¹ L ² AFRAAHRYLTAV	1368
1250	Db	DLPADHARPMPRSPFAGDKVSFTLDQEVTSGLYKLARENGSTLYNWLAA ¹ YATFSLRSLGQ	1309
1369	QY	EDAVIGTPIANRN ¹ PELEDIIIGCEVNTQC ¹ MRINIDHDHDTGTGLINQVYKATTTAA ¹ FENEDI	1428

```
Db 1310 EDIIVGSIAGRPKHUDEPIILGMEVNTIALTRPEGKPFVQYLOEVRETAMEAFERODY 1369
QY 1429 PFERVVSALOPGSRDLSSTPLAQILFAVHSQKDLGRKFQGLSEVPVPSKAY--TRFDEME 1486
Db 1370 PFEELVOKLEL-TRDMSRNPFLVDMVFLQNM-DOESLELDELCLKPAANNHOSHKSDEL 1427
QY 1487 FHLFOETDS-LKGSVNFADLFKMETVENVVRVFEILNGLQSSRTPVSLPTDGI 1545
Db 1428 LYAQEOGRLLTFQMEFSTDLKTKIEKWLQYENNNLLSIKDKNAALGTINILNEDEA 1487
QY 1546 LEKLDVLNVHVDYPRESSLADVPOTVSAYPDSLAVVDSCLTYELTDROSILAGWL 1605
Db 1488 HYLHELNRKIDIPRNETISRLPEMQAEQTPNAVAIVSDQVFTYEDLNSWANQIASVL 1547
QY 1606 RRRSMPAETLVAVPAPRSCETIVAFFGVGLKANLAYPLDVRPSGARVQDILSGSGPTIV 1665
Db 1548 QIKGVGPDVVALLTGRTPELIAGMLGILKAGGAYLPIDSNLPIVERTAYMLS--DSRAAL 1605
QY 1666 LIGHDTAPPDIEVNVFEVRIRDALNDSNADGFEVIEDHSTKPSATSLAVVLYTSGSTGR 1725
Db 1606 LLOSEKTEKRLUGIECEQIIIEDIQKQGEAKNVE----SSAGPH--SLAYIITYSGSTGK 1659
QY 1726 PKGVMIHRVIRIVTSCIPNYPSETRMAHMATIAFDGASYEYISALLFCRTLVCVDYM 1785
Db 1660 PKGVMIORSVIRLVKNSNITFTPEDKLLMTSSIGFDVGSFEIFGELLNGAALHLSDOQ 1719
QY 1786 TTLDRALKDVFFREHVNASHVTSQDQVPLRPRLSLTLMFFLVVTDSTAPDALD- 1844
Db 1720 TFLDSHOLKR--YIEHQGITIMLTSLFNHLTEQNEOTFSQLKHLIIGCEALSPSHNR 1777
QY 1845 AQLYQGVQCYNGVGTENGVMSTIYIDSTESFINGVPIGRALNLSGAYVVDPEQOLVG 1904
Db 1778 IRNVCPEVSNWNGVGTENTFTFSLHIQKTYEL--SIPICRPVGNSTAFILNOMGVLP 1835
QY 1905 IGVMGELVGTGDLARGYSKA-IDENRFV--HITVNDQTVKAYRTGDRVRYRIGDGLIE 1961
Db 1836 VGAVGELCVGDGVARGYLGRPLDTEKFPHPFAPGD--RLYRTGDLARW-LSOGTIE 1891
QY 1962 PFGMDTQFKIRGNRIESAIEAALLRDSVVRDAVVLQONEDQAEILGFVWADHDHSE 2021
Db 1892 YVGRIDQVQVRYGVELGETALRQIDGVEAAVLARTAQTKGSKELFGYISV----- 1945
QY 2022 NDKQSANOVE-----GWODHFECSMYSDIGEIDPSTIGSDPKGWTSMYDGS 2068
Db 1946 -KAGTNAEQVRLARSLLPNMIPAYIEMETPLTNSGNLKRKALPEPDVASKQTYIPP 2004
QY 2069 QIDFDEM-----HEWLGETTRTLHNRSLGNVLEIGTSGMILFNLDLSRESY-VGLEPS 2122
Db 2005 RNELEEQALIQEVLGIQRIQIED-----SFFELG-GDSIKALQVSARLGRYGLSLOVS 2058
QY 2123 -----RSAAAFVNKATESIPSILAGKAKVQVGTATDIGO-----VDDLHPDLVVLNS 2168
Db 2059 DLFRHPRIKIDLPFIRKSERIIEQ--GPIQGDV--PWTVPQWFFSQDIEERHH-----FNQ 2111
QY 2169 VIQVFE-----PSSEVLAEIADTLIHLPNVQRIFFGDVR-----SQATNEHFLAA 2212
Db 2112 SVMLFHSGRLSENALRPALKLAHHHDAL-----RMVYRNDRRWTOINOIHESOLYS 2165
QY 2213 RAIHTLGKNAT--KDDVRQKMAELE---DMEELLVEPAPFTSLKDRFPGLVEH----- 2261
Db 2166 LRISDLQSSEGWETKIQEVADLQOSINLOEGPLLHAALFKTLTG DYLF LAIHLVVDG 2225
QY 2262 --VEILPKMNEAVNELSAYRYAAVHVVRGSLGDELVLPEKDDWIDFQANQLNQKSLGDL 2319
Db 2226 VSWRILLELSA-----GYQOAAA-----GOTIQLPKPTDSYQY-ARRIOEYAOSSK 2272
QY 2320 LKSSDAIMAV-----SKIPFEITAFERQVAVSINSNIDEWQLSTIR--SSAEGDSSLSVP 2373
Db 2273 LIREAYWRSVEEQAAELPYEIPH-----HVNIDFSKRDSLSLSLSTEADTAVLLQ 2323
QY 2374 DIFRIAGEAGFRVEVSSAR-----QWSQNGALDAVFHCCSQGR----- 2412
```

2324 NVNHAYGTDTODIILLTASLAICEWGTGSGKILRIAME---GHGREHILPELDISRTVGWFT 2380

2413 ----TLVNFTDHHLRGSDLLTNR-PLORLQNRRIAEVRRERLRSLLPSYMIPSN-IVVL 2466

2381 SMYPALISFENHRDELGTSVKTVKDTLCRIPNKGVGY-----GMLKYLTHPENKSTTF 2433

2467 DKMP-LNANGVKDRKELSSRAKVVPKQOATAAPLPTFPPISEVEVILCEATEVFGMKVDIT 2525

2434 SKTPEISFENYLGQFNDIERQDTERPSSILSGSKDITHTWKREQII-----EWSAMAADKK 2487

2526 DHFFNLG-----GHSLLATKILSRIDQRL-----KVRITVKDVFDPHPFA-DLA 2568

2488 LH-FNLSYPPARFRHNTMEQLINRIEHFLDIMKHCAGQQAECTLSDFSQSLTAEDLD 2546

2569 SV 2570

2547 SI 2548

Search completed: May 30, 2003, 12:50:02

Job time : 164 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protéin - protein search, using sw model

Run on: May 30, 2003, 12:39:11 ; Search time 161 seconds
(without alignments)
4004.481 Million cell updates/sec

Title: US-09-482-788-2
Perfect score: 16128
Sequence: 1 MEYLTAVDGRQLPPTPASF.....RVEHLLLEVSKTPEGLNSSL 3129

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bactexiap.*
- 17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9874	61.2	3131	3 Q00869	Q00869 fusarium eq
2	5000	31.0	15281	3 Q09164	Q09164 tolypocladi
3	2871	17.8	983	3 Q00868	Q00868 gibberella
4	2496	15.5	1051	3 P97961	P97961 cylindrotri
5	2391	14.8	4247	2 Q9L8H4	Q9L8H4 streptomyc
6	2275.5	14.1	4848	2 Q07944	Q07944 streptomyc
7	2146.5	13.3	2841	2 Q9FB33	Q9FB33 streptomyc
8	2108	13.1	5060	2 Q9K5M1	Q9K5M1 anabaena sp
9	2018.5	12.5	5157	3 Q01135	Q01135 metathizium
10	2002	12.4	4360	3 Q9UVN5	Q9UVN5 alternaria
11	1984	12.3	1997	2 Q05647	Q05647 streptomyc
12	1978.5	12.3	3317	2 Q9RAH2	Q9RAH2 nostoc sp.
13	1974	12.2	4379	2 Q9RAH4	Q9RAH4 nostoc sp.
14	1933	12.0	4450	2 Q44928	Q44928 bacillus br
15	1906.5	11.8	3670	16 Q924X5	Q924X5 streptomyc
16	1896.5	11.8	2588	16 Q8YTR5	Q8YTR5 anabaena sp

17	1892.5	11.7	6889	16 Q8XS40	Q8XS40 ralstonia s
18	1880.5	11.7	2450	2 Q9RAH1	Q9RAH1 nostoc sp.
19	1879	11.7	5953	16 Q8XS39	Q8XS39 ralstonia s
20	1876.5	11.6	9376	2 Q85168	Q85168 pseudomonas
21	1867.5	11.6	5149	16 Q91179	Q91179 pseudomonas
22	1866.5	11.6	3588	2 Q66070	Q66070 bacillus li
23	1860	11.5	3310	2 Q9AMR5	Q9AMR5 bradyrhizob
24	1831.5	11.4	3583	2 Q45675	Q45675 bacillus su
25	1828	11.3	2258	2 Q9K5M2	Q9K5M2 anabaena sp
26	1811	11.2	3589	2 Q69246	Q69246 bacillus li
27	1810	11.2	7463	16 Q924X6	Q924X6 streptomyc
28	1803	11.2	2554	2 Q30981	Q30981 bacillus su
29	1795	11.1	3316	2 Q8VQF8	Q8VQF8 xenorhabdus
30	1781.5	11.0	2617	16 Q8YTR9	Q8YTR9 anabaena sp
31	1755	10.9	2571	2 Q87704	Q87704 bacillus su
32	1729.5	10.7	5369	2 Q9R9J0	Q9R9J0 bacillus su
33	1725	10.7	3603	16 P94459	P94459 bacillus su
34	1715	10.6	4898	2 Q93N87	Q93N87 streptomyc
35	1713.5	10.6	2448	16 Q91182	Q91182 pseudomonas
36	1713	10.6	5362	2 Q93155	Q93155 bacillus su
37	1708.5	10.6	2448	2 Q51338	Q51338 pseudomonas
38	1686.5	10.5	2246	2 Q9AKS6	Q9AKS6 pseudomonas
39	1677.5	10.4	3582	2 Q66059	Q66059 bacillus li
40	1677	10.4	4342	16 Q91157	Q91157 pseudomonas
41	1672	10.4	2611	2 Q68487	Q68487 streptomyc
42	1649.5	10.2	2378	2 Q9R912	Q9R912 bacillus su
43	1634.5	10.1	499	3 Q96UG8	Q96UG8 fusarium pa
44	1629.5	10.1	3232	3 Q94205	Q94205 claviceps p
45	1625	10.1	3583	2 Q45295	Q45295 bacillus li

ALIGNMENTS

RESULT 1

Q00869	ID	Q00869	PRELIMINARY;	PRT;	3131 AA.
AC	Q00869;				
DT	01-NOV-1996 (TREMBLrel. 01, Created)				
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)				
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)				
DE	Eniatiin sythetase.				
GN	ESYN1.				
OS	Fusarium equiseti.				
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;				
OC	Hypocreales; mitosporic Hypocreales; Fusarium.				
OX	NCBI_TaxID=61235;				
RP	[1]				
RC	SEQUENCE FROM N.A.				
RC	STRAIN=LAMBOTTE ET FAUTREY;				
RC	MEDLINE=93247491; PubMed=8483420;				
RA	Haese A., Schubert M., Herrmann M., Zocher R.;				
RT	"Molecular characterization of the euaitiu synthetase gene encoding a				
RT	multifunctional enzymecatalysing n-methyl depsiptide forluation in				
RT	Fusarium scirpi.;"				
RL	Mol. Microbiol. 7:905-914(1993).				
RN	[2]				
RC	SEQUENCE FROM N.A.				
RC	STRAIN=LAMBOTTE ET FAUTREY;				
RA	Zocher R.;				
RN	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.				
RC	[3]				
RC	SEQUENCE FROM N.A.				
RC	STRAIN=LAMBOTTE ET FAUTREY;				
RA	Zocher R.;				
RL	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; Z18755; CAA79245.2; -				
DR	HSSP; P14687; IAMU.				
DR	InterPro; IPR000873; AMP-bind.				
DR	InterPro; IPR001242; Condensatn.				
DR	InterPro; IPR003880; Prantne_attach.				
DR	Pfam; PF00501; AMP-binding; 2.				
DR	Pfam; PF00668; Condensation; 2.				

DR Pfam: PF00550; pp-binding: 3.
DR PROSITE: PS00075; ACP_DOMAIN; 6.
DR PROSITE: PS00455; AMP_BINDING; 2.
DR PROSITE: PS00112; PHOSPHOPANTHETHEINE; UNKNOWN_3.
KW Phosphopantetheine.
SQ SEQUENCE 3131 AA; 346494 MW; AD7663E91FAB67C4 CRC64;

Query Match
Best Local Similarity 61.2%; Score 9874; DB 3; Length 3131;
Matches 1959; Conservative 448; Mismatches 680; Indels 86; Gaps 30;

QY 1 MEYLTVADGRODLPPTPASCSHGSDPLNSSYEOFLHYGLDLSRIEAIKPCPTFQDLM 60
DB 1 MSLHTPSGGQDPALASKTLC-----EQLSALGLGQDKIENIFPGTFQRDVI 49
QY 61 DCAALDRQSAIGHAVYDPTDISREFALEAWKEVNTGTPALRAFAFTSDSKTKTSQVILKD 120
DB 50 DCAADKQRAVGHAVEIPKIDDAARLAANKETVLTPTALRCTFTTSKSGDVLQVLRD 109
QY 121 SFVSWCWSSSSPDEVVDEAAAAAGPRCNREVLLDMQTKKCOLVWTFSHALVDVT 180
DB 110 SFVSWSGSPVDLKEAVQVDEAAALAGPRCNREVLLDPDKERQLIWTFSHALVDST 169
QY 181 FOORVLSRVAAYKHEDTH--REPETSSDATD-----SQSVSVNSCEDNAVSA 232
DB 170 FOERILRRVLYKADANDEHPQETPDSSQATPEEDLQPNPSKMLKIPQAADMRAV-- 227
QY 233 THWQTHLDNLNASVFFPHSLDHLMVNPNTTAAEHRTFFPLS-QKALSNSAICRTALSILL 291
DB 228 -EFWKDHLGLKCFCLPAFVLSVYAPHDAKAEHRISYSSAQKMSATICTALAILL 286
QY 292 SRYTHSDEALFAGTEQSLPDKHYLDAGYQTVAPLRVHCOSNLRASDWDMAISSYDDR 351
DB 287 SRYTHSPALFVIGTEQTLLEQMLDGPTRVTPVIRVSCASEQSVSTIMSTISYDQT 346
QY 352 LGHLAPGLDRIRNTGNSAACDFQTVLLVTDGSHVNNINGFLQOITESSHFMPCNNR 411
DB 347 MRQFAHAGLRNIASAGDESAACGFQVLLVSDGDAQPASTWEILLKATEPEGEPICTNR 406
QY 412 ALLLHCQESSGALLVAYDHNVIDSLQTRLLQFQGHLLKCLQSPDLDSMAEVLNMT 471
DB 407 ALLLSCQMTSSGAHLTARYDQSIIDAQEMARLLRQLGHLIQLNQTSTDL-PVEKVDMMTQ 465
QY 472 YDRAIESWNSQPLEVQDTHLHHEMLKAVSHSPKTAIQAWDGDWNTSELDNVSSRLAVH 531
DB 466 EDWLEIERWNSDSIDAQDTLHSEMKTWTSQSPNKAATAVADGEMTWYAEILDNVSSRLAQH 525
QY 532 IKSLGLRAQAIIPVPEKSKWVIASMLAVLKSGNAFTLIDPNDPPARTAOVVTQTRATV 591
DB 526 INSIDLCKEHAIVPIYEKSKWVAMSLAVLKAGHAFTLIDPDPARTAOVVOQTSATV 585
QY 592 ALTSKLRHRETQKLVGRVVDDELQSVSASDDFSLTKSQDLAYVIFTSGSTGDPKGI 651
DB 586 ALTSKLRHRETQSVTVGRVIVDEFEVKSLPQSSSELSASVKAHLAYVIFTSGSTGIPKGI 645
QY 652 MIEHRFASSCALAFGASLGINSOTRALQFGTHAFGACLLIEMTTLINGCVCIPSDDDRM 711
DB 646 MIEHRFASSCAIFGALPGLTSTRALQFGSHAFGACILIEMTTLIHGGCVCIPSDDDRM 705
QY 712 NSIPSFNRYNNVMNMTAPSYMGTFSPEDVPGLATLVLVEQMSSSVNAIWAQKLQLLNG 771
DB 706 NNVLFINRTNVQLGHATPSYMGTFQEVVPGKLTVLVGEQMSASVNEWAPVQLLNG 765
QY 772 YGSESSSICFASNM-----STEPNNMGRVAGHNSWIDPNDINRLVPIGAVGELVIESPGI 828
DB 766 YGSESSSICCAKISPCSSEPNIGHAVGHSNIVDPEDPNRLAPIGAVGELVIESAGI 825
QY 829 ARDIYVPPPEKSPFFTDISWPANTFPDGAKLYRTGDLARVASDGSIVCLGRIDSQVK 888
DB 826 ARDIYVAPTQDKGPFITKAPTAPKQLPDGFIYRTGDLACVASDGSIVCLGRMDSQVK 885
QY 889 IRGORVELGAIETHLRQOMDDTLTVVEATKRSQSANSTSLIAFLIGSSYFGNRPSPDAHI 948
DB 889 IRGORVELGAVETHLRQOMDDTLTVVEAVKFSFSSSTTVLTAFLIGAG-----EKNSHI 940

DB 886 IRGORVELGAVETHLRQOMDDTLTVVEAVKFSFSSSTTVLTAFLIGAG-----EKNSHI 940
QY 949 LDHATKAINIKLEQVLPRHSIPSFYICMLERPTATCKIDRRRLRMKIDILQKQTOCA 1008
DB 941 LDQRATREINAKMEQVLPRHSIPAFYISNMNLQPTATCKVDRLRIMSKILSQKTHST 1000
QY 1009 IVQAPAPIPVFAADTAAKLHSIWQSLGIDPATVNVGATFPFELGNSITAKIMVNMAHSV 1068
DB 1001 PSQSQAAISSGTDYTKLESITWITSLDLEPGSANMSATFFEMGNSIIAIKIMVNMAHSN 1060
QY 1069 GMDLKNVNIYQHPHTLAGISAVVKGDPISYTLIPLKSTHEGPV-EQSYSGRKLWFLDOLDVG 1127
DB 1061 GIELVSDIYQNPHTLAGIKAIVGTSPLIPKVTQGPVSEQSYAONRMWFLDQSE 1120
QY 1128 SLWYLIPIYVYMRGPNVNDALRRALAALAEQORHETLRTTFEDQDGVQVIVHEKLSSEMKV 1187
DB 1121 ASWYLIPIYVYMRGPNVDALTRALLALEQORHETLRTTFENQDGVQVIVHRLSKELQV 1180
QY 1188 IDLCGSDLDPEVLNQEQTPFPNLSSEAGWRATLLRGEDDHILITVIMHHIISDGSIDV 1247
DB 1181 IDALDGGDEGLTKLYKVETTFDITSEAGWSSTLIRLGKDDHILSIIMHHIISDGSIDV 1240
QY 1248 LRDLNOLYSAALKSDKPLSALTPLPTIOYSDFAKWOKD--QFTEQEKQLANWKKOLKDS 1305
DB 1241 LRRELQIYAAALQ-GKDPSSALTPLPQISDFAVWQKQEAQAHEHQLOIYWKQLADS 1299
QY 1306 SPARKIPTDFARPALLSGDAGCVHYTIDGELYQSLRAFACNEHNTTSFVYLLAAFAAAHYRL 1365
DB 1300 SPARKIPTDFPRDLLSGDAGVVPVADGELYQKLRGCKNKHNSSTAFSILLAAFAAAHYRL 1359
QY 1366 TAVEDAVIGTPIANRNPPELIDITGCFVNTQCMRINDHDTFOTGLINQVATTAAPE 1425
DB 1360 TAVDDAVIGTPIANRNRWELENMIGFEVNTQCMRIADVEDTTFESLRQVSTTTAAFAH 1419
QY 1426 EDIPFERVWSALQSGRDLSTPLAQLIFAVHSOKDLGRPKFGQESLPVPSKAYTRFDM 1485
DB 1420 EDVFERVWSALQGHRLSRTPLAQIMFAVHSOKDLGRFELEGIQSEPIASKAYTRFDV 1479
QY 1486 EFHLFOETSLDKGSVNFADLEFKMETVENVYVFEILRNLQSSRTPVSILPLTDGIVT 1545
DB 1480 EFHLFOQADGLKGCNEFATDLFKPETIIONVYVVFQILRHGLDQPETCISVPLTDGVEE 1539
QY 1546 LEKLDVLNVKHVDYPPRESSLADVPQTOVSAYPDSLAVVDDSCRLTYTIELDQSDILAGWL 1605
DB 1540 LRRDLLEIKRTNYPDRSSVVDVFEQAAANPEVIAVTDSSRLTYAELDNKSELLRWL 1599
QY 1606 RRRSMPAETLVAVFAPRSCETIVAFVGLKANLAYPLDVRSPSARVQDILSGLSGPTIV 1665
DB 1600 RRRNLTPETLVSLVAPRSCETIVAYVGLKANLAYPLDVRSPVTRMKDILSSVSGNTIV 1659
QY 1666 LIGHDTAPDIEVTNVFEVIRDALNSNADGFVEIHDSTKPSATSLAYVLYTSGSTGR 1725
DB 1660 LMGSVDEPDGFDLPQLELVRITDFTDETIED-----VQDSPQSATSLEYVFTSGSTGK 1714
QY 1726 PKGVMIEHRLVIRVTSCIPNYPSETRMAHMATIAFDGASYEYISALLFGRTLVCVDYM 1785
DB 1715 PKGVMIEHRAIVRLVKSDFNPGFPPSPARMSNVFNPAFDGAIWEINMLLNGTVVCIDYL 1774
QY 1786 TTLDARALKDVFREHVNAASHVTSSSQDVPRLVRPRLSRSLTMEFFLVVTDSTAPDALDA 1845
DB 1775 TTLDGKELAAVFAKERYNAFAFAPAMLK-LYLVDAAREALKNLDFLIVGGERFDTKEAVEA 1833
QY 1846 QGLYQGVQCYNGVPTENGVNSTIYPIDSTESFINGVPIGRALNNSGAYVVDPEQLVGI 1905
DB 1834 MPLVRG-KIANIYGPTEAGIISTCYNIPKDEAVTNGVPIGGSYVNSGAYVMDPNQQLVGL 1892
QY 1906 GVWGLVVTGDGLARGYSKALDENRVHTVNDQTKAYTGRDVRVYRIGDGLIEFGR 1965
DB 1893 GVWGLVVTGDGVRGYTNPELNKNRFDITIEGKTFKAYTGRDVRVYRIGDGLIEFGR 1952
QY 1966 MDQFKIRGNRIESAETEAALLRDSRVDAAVLIQ-QNEQAP-EILGLFVVADHDHSDND 2023
DB 1953 MDNOKIRGNRIEAGEVESAMLSLKNVNLAAIVVRGGDEBGPVLEWGVFIVADDKNDTTE 2012

Qy	2024	KG	S	A	N	O	V	E	G	W	Q	H	F	E	S	G	M	Y	S	O	I	G	-	E	I	D	P	S	T	I	G	S	D	F	K	G	W	T	S	M	Y	D	G	S	O	I	D	E	M	H	E	M	L	G	E	T	2082			
Db	2013	E	E	E	T	G	N	O	V	E	G	W	Q	H	F	E	S	G	M	Y	S	D	I	S	T	A	V	D	Q	S	A	I	G	N	D	F	K	G	W	T	S	M	Y	D	G	K	I	D	K	G	E	M	Q	E	W	L	D	D	A	2072
Qy	2083	T	R	L	H	N	R	S	L	G	N	V	L	E	I	C	T	G	S	M	I	L	F	N	L	D	S	R	E	S	V	G	L	E	P	S	R	S	A	A	F	N	K	A	T	E	S	I	P	S	L	A	C	K	2142					
Db	2073	I	H	T	L	H	N	G	O	I	P	R	O	V	L	E	I	C	T	G	S	M	I	L	F	N	L	P	N	L	G	N	S	V	G	L	D	S	K	S	A	V	E	F	N	R	A	V	E	S	P	K	F	A	G	2132				
Qy	2143	A	K	V	O	G	T	A	D	I	G	O	V	D	L	H	P	L	V	V	L	N	S	I	O	I	Y	F	P	S	E	Y	L	A	E	T	A	D	T	L	I	H	L	P	N	V	O	R	I	F	F	G	D	V	R	S	2202			
Db	2133	A	K	V	H	G	M	A	T	D	V	N	K	L	G	E	V	H	P	D	L	V	F	N	S	V	O	Y	F	P	T	P	E	Y	L	A	E	V	I	D	G	L	I	A	I	P	S	V	K	R	I	F	L	D	I	R	S	2192		
Qy	2203	Q	A	T	N	E	H	F	L	A	A	R	I	H	T	L	G	--	K	N	A	T	K	D	D	Y	R	O	K	M	A	E	L	D	E	M	E	E	L	L	V	E	P	A	F	F	T	S	L	K	D	R	F	F	L	C	V	E	2260	
Db	2193	Y	A	T	N	G	H	F	L	A	A	R	I	H	T	L	G	T	N	N	A	T	K	D	R	O	K	I	Q	E	L	D	E	R	E	E	F	L	V	E	P	A	F	F	T	T	L	K	E	R	R	P	D	V	V	K	2252			
Qy	2261	H	V	E	I	L	P	K	N	M	E	A	V	E	L	S	A	R	Y	A	A	V	V	H	V	R	G	S	L	G	D	E	L	V	P	E	K	D	D	I	D	C	A	N	O	L	N	O	K	S	L	G	D	L	L	2320				
Db	2253	H	V	E	I	L	P	K	N	M	A	T	E	L	S	A	R	Y	A	V	V	H	L	R	D	E	T	--	D	E	P	V	I	H	I	E	K	D	S	V	D	E	F	A	K	O	M	D	T	A	L	L	D	H	L	2311				
Qy	2321	K	--	S	S	D	A	I	M	A	V	S	K	I	P	E	I	T	A	F	R	O	V	A	S	L	N	S	I	D	E	-----	W	O	L	S	I	R	S	S	A	E	G	D	S	2369														
Db	2312	R	L	S	K	D	A	M	S	A	V	S	N	I	T	A	H	A	F	E	R	I	V	E	S	L	D	E	S	K	D	O	T	K	G	L	D	C	A	A	W	--	L	S	A	V	R	E	A	E	N	R	A	S	2370					
Qy	2370	L	S	V	P	I	R	I	A	G	E	A	G	R	V	E	S	S	A	R	O	M	G	A	L	D	A	V	F	H	I	C	--	C	S	O	G	R	T	L	V	N	P	E	T	D	H	L	G	S	D	2427								
Db	2371	L	T	V	P	O	L	E	I	A	K	E	A	G	R	V	E	S	A	R	O	M	G	S	G	A	L	D	A	V	F	H	I	P	P	S	T	D	R	L	I	O	F	P	T	O	N	E	L	R	S	S	L	2430						
Qy	2428	L	L	N	R	P	L	O	R	L	O	N	R	R	I	A	E	V	R	L	R	S	L	P	S	M	I	D	S	N	I	V	L	D	K	M	P	L	N	A	K	V	D	R	E	L	S	R	R	A	K	2487								
Db	2431	T	L	N	R	P	L	O	K	L	O	R	R	A	L	O	V	E	K	I	O	T	L	V	P	S	M	V	P	N	I	V	L	D	T	M	P	L	N	T	N	G	K	I	D	R	E	L	T	R	R	A	2490							
Qy	2488	V	P	K	O	T	A	A	P	L	P	T	P	I	S	E	V	E	V	I	C	E	A	T	E	V	F	G	M	K	V	D	I	T	H	F	N	L	G	H	S	L	A	T	K	L	S	I	D	Q	2547									
Db	2491	T	L	P	K	O	T	A	A	P	V	P	D	T	S	I	D	E	I	T	C	E	A	T	E	V	F	G	M	K	V	E	L	S	D	H	F	F	O	L	G	H	S	L	A	T	K	L	S	I	Q	H	2550							
Qy	2548	R	L	K	V	R	I	T	V	K	D	V	P	D	H	P	F	A	L	S	V	I	R	O	G	L	Q	P	V	S	D	G	Q	--	O	D	R	S	A	H	M	A	P	E	T	A	I	L	C	D	E	2606								
Db	2551	R	L	H	V	R	V	T	V	K	D	V	P	S	P	F	A	D	L	A	V	I	R	O	G	L	M	O	N	P	V	A	E	G	S	O	D	K	O	G	N	S	S	R	V	A	P	R	T	E	V	E	K	M	L	C	E	2610		
Qy	2607	F	A	K	V	L	G	F	O	V	G	I	T	N	F	D	L	G	H	S	L	M	A	T	K	L	A	V	R	I	G	H	R	L	D	T	T	V	S	V	K	D	V	D	H	-----	P	V	L	F	Q	2659								
Db	2611	F	A	A	G	L	V	P	V	G	I	T	N	F	D	L	G	H	S	L	M	A	T	K	L	A	V	R	I	G	R	L	-----	I	R	H	S	O	G	H	L	R	P	C	A	F	Q	2662												
Qy	2660	L	A	T	A	D	N	L	V	O	S	K	T	E	N	I	V	G	R	E	M	A	E	S	P	F	O	L	L	T	E	D	P	E	E	M	A	S	E	T	K	P	O	L	E	--	L	Q	E	I	O	D	I	Y	P	2718				
Db	2663	L	A	K	L	E	S	H	S	K	E	S	E	S	D	O	L	M	A	D	Y	A	F	L	D	L	D	E	D	P	O	F	V	O	S	I	R	P	O	L	D	S	C	Y	T	I	Q	D	V	I	P	2722								
Qy	2719	S	T	O	M	O	K	A	F	L	D	H	T	T	A	R	P	R	P	V	E	P	I	D	P	P	S	T	E	P	D	A	A	G	L	I	K	A	C	E	S	L	N	H	L	D	I	F	R	T	V	A	E	A	S	2778				
Db	2723	S	T	O	M	O	K	A	F	L	D	P	T	T	G	E	R	G	L	V	E	I	D	P	S	N	A	--	D	A	E	T	L	T	K	A	I	G	A	L	V	D	K	L	D	M	F	R	V	F	L	E	A	2780						
Qy	2779	G	E	L	Y	Q	V	V	L	S	C	L	D	I	P	I	Q	V	I	E	T	E	D	N	T	A	T	N	E	L	D	E	F	A	K	E	P	V	R	L	G	H	P	L	I	R	T	I	I	K	O	T	K	S	M	R	2838			
Db	2781	G	D	L	Y	Q	V	V	E	H	N	L	P	I	E	T	E	K	N	V	N	T	A	T	G	D	Y	L	D	V	H	G	K	D	P	V	R	L	G	H	P	C	I	O	F	A	I	L	K	T	A	S	S	V	2840					
Qy	2839	I	M	R	I	S	H	A	L	D	G	L	S	L	E	H	V	R	K	L	M	L	N	G	R	S	L	L	P	H	O	F	S	R	Y	M	O	Y	T	A	D	G	R	E	S	H	G	E	W	R	D	I	O	N	2898					
Db	2841	L	L	R	M	S	H	A	L	D	G	L	S	F	E	V	I	R	G	L	V	I	S	G	R	N	L	P	P	P	T	O	F	A	R	Y	M	O	Y	A	A	H	S	R	E	E	G	Y	P	M	W	E	V	I	Q	N	2900			
Qy	2899	T	P	M	T	I	L	S	D	--	D	T	V	V	D	G	N	A	T	C	K	A	L	H	L	S	K	I	N	I	P	S	O	V	L	R	G	S	S	N	I	T	O	A	T	V	E	N	A	A	C	A	L	V	L	S	R	2956		
Db	2901	A	P	M	T	V	L	D	T	N	N	G	S	E	O	E	M	P	A	S	K	A	V	H	L	S	E	V	N	V	P	A	I	R	N	S	T	N	--	T	O	A	T	V	E	N	A	C	A	L	V	L	A	K	2958					
Qy	2957	E	S	O	S	K	D	V	V	F	G	R	I	S	V	G	R	O	L	P	V	E	O	I	V	G	P	C	T	N	A	V	P	V	R	A	H	I	E	S	S	D	N	O	L	L	H	D	I	O	D	Y	L	L	S	3016				
Db	2959	E	S	G	S	D	V	V	F	G	R	I	S	V	G	R	O	L	P	V	W	O	D	I	I	G	P	C	T	N	A	V	P	V	H	A	R	V	D	G	N	P	O	I	I	R	D	L	R	D	Y	L	R	T	L	3018				
Qy	3017	P	H	E	T	I	G	S	D	L	K	R	N	C	T	D	H	P	E	A	I	T	N	F	S	C	C	I	T	H	N	E	F	Y	H	P	E	S	O	F	E	O	R	V	E	M	G	V	L	T	F	V	N	E	M	D	3076			
Db	3019	P	F	E	S	L	G	P	E	E	L	K	R	N	C	T	D	M	P	E	E	L	T	N	F	S	C	V	T	Y	H	N	E	F	Y	H	P	E	S	E	V	D	N	O	K	E	M	G	V	L	A	K	V	E	L	S	E	N	3078	

Qy	3077	EPLYDLAIAAGEVDPDGAAGLKVTVIATQIFGRRKRVEHLLFVSKTFEGUNSSL	3129
Db	3079	EPLYDLAIAAGEADGVNLKVTVAARLYNENRHRHVEEYCKTFNGLINEAL	3131
RESULT 2			
ID	Q09164	PRELIMINARY;	PRT; 15281 AA.
AC	Q09164;		
DT	01-NOV-1996	(T-EMBLrel. 01, Created)	
DF	01-NOV-1996	(T-EMBLrel. 01, Last sequence update)	
DT	01-MAR-2002	(T-EMBLrel. 20, Last annotation update)	
DE	Cyclosporin synthetase (CYSYN)	(EC 6.-.-.-)	
OS	SIMA.		
GN	Tolypocladium inflatum.		
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;		
OC	Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae;		
OC	Tolypocladium.		
OX	NCBI_TaxID=29910;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.		
RC	STRAIN=ATCC 34921;		
RX	MEDLINE=95094306; PubMed=8001164;		
RA	Weber G., Schoerndorfer K., Schneider-Scherzer E., Leitner E.;		
RT	The peptide synthetase catalyzing cyclosporine production in		
RT	tolypocladium niveum is encoded by a giant 45.8-kilobase open reading		
RT	frame."		
RL	Curr. Genet. 26:120-125(1994).		
CC	-1- FUNCTION: THE CONSTITUENT AMINO ACIDS OF CYCLOSPORINS ARE		
CC	ACTIVATED AS AMONACHYL-ADENYLATES WITH PEPTIDE BONDS FORMED		
CC	THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER INTERMEDIATES.		
CC	ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.		
CC	-1- COFACTOR: CONTAINS 11 COVALENTLY BOUND PHOSPHOPANTHEINES.		
CC	-1- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF CYCLOSPORINS.		
CC	-1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT		
CC	COVALENT BINDING OF AMP TO THEIR SUBSTRATE.		
CC	EMBL; Z28383; CAA82227.1; -.		
DR	HSSP; P14687; IAMU.		
DR	InterPro; IPR000873; AMP-bind.		
DR	InterPro; IPR000267; Asp/Glutamase.		
DR	InterPro; IPR001242; Condensatn.		
DR	InterPro; IPR003880; Ppantne.attach.		
DR	InterPro; IPR000051; SAM_bind.		
DR	Pfam; PF00501; AMP-binding; 11.		
DR	Pfam; PF00668; Condensation; 13.		
DR	Pfam; PF00550; pp-binding; 11.		
DR	PRINTS; PR00154; AMPBINDING.		
DR	PRINTS; PR00139; ASGNLASE.		
DR	PROSITE; PS00075; ACP_DOMAIN; 11.		
DR	PROSITE; PS00455; AMP_BINDING; 10.		
DR	PROSITE; PS0012; PHOSPHOPANTHEINE; UNKNOWN 9.		
KW	Ligase: Antibiotic biosynthesis; Phosphopantetheine; repeat;		
KW	Multifunctional enzyme.		
FT	DOMAIN 15179 15219	13 X 3 AA APPROXIMATE REPEATS.	
FT	REPEAT 3 1086	DOMAIN 1.	
FT	REPEAT 1087 2585	DOMAIN 2.	
FT	REPEAT 2586 4072	DOMAIN 3.	
FT	REPEAT 4073 5564	DOMAIN 4.	
FT	REPEAT 5565 7061	DOMAIN 5.	
FT	REPEAT 7062 8121	DOMAIN 6.	
FT	REPEAT 8122 9616	DOMAIN 7.	
FT	REPEAT 9617 11113	DOMAIN 8.	
FT	REPEAT 11114 12185	DOMAIN 9.	
FT	REPEAT 12186 13681	DOMAIN 10.	
FT	REPEAT 13682 14767	DOMAIN 11 (ALA-ACTIVATING).	
FT	BINDING 1060 1060	PHOSPHOPANTHEINE (POTENTIAL).	
FT	BINDING 2558 2558	PHOSPHOPANTHEINE (POTENTIAL).	
FT	BINDING 4045 4045	PHOSPHOPANTHEINE (POTENTIAL).	
FT	BINDING 5537 5537	PHOSPHOPANTHEINE (POTENTIAL).	
FT	BINDING 7034 7034	PHOSPHOPANTHEINE (POTENTIAL).	
FT	BINDING 8094 8094	PHOSPHOPANTHEINE (POTENTIAL).	
FT	BINDING 9589 9589	PHOSPHOPANTHEINE (POTENTIAL).	
FT	BINDING 11086 11086	PHOSPHOPANTHEINE (POTENTIAL).	

FT BINDING 12158 PHOSPHOPANTHETHEINE (POTENTIAL).
FT BINDING 13654 13654 PHOSPHOPANTHETHEINE (POTENTIAL).
FT BINDING 14729 14729 PHOSPHOPANTHETHEINE (POTENTIAL).
SQ SEQUENCE 15281 AA: 1689051 MW: E26DA7AA35324C05 CRC64;
Query Match 31.0%; Score 5000; DB 3; Length 15281;
Best Local Similarity 32.4%; Pred. No. 1.2e-299;
Matches 1233; Conservative 440; Mismatches 967; Indels 1264; Gaps 66;
QY 482 SOPLEVQDTLIHHEMLKAVSHSPTKTAIOAWDGDWYSSELDNVSRLAVHIKSLGLRAOQ 541
DB 11581 NQPDYPRDASVIDVFPQVASIPKSIATAVIDASSQLTYTDLDRSSQLATWLR-QVTYPE 11639
QY 542 AIIPVYFEKSWVIASMLAVLKGNAFTLIDPNDPPARTAQVYTOR--RATVALTS----- 595
DB 11640 ELGVGLAPRSCETIIAFLGIKLANLAYPLDWNAPAGRIETILSSLPNGRLIILGSDTQA 11699
QY 596 -KLHRETQVKGRCVVVDDELLOSYS-ASDDFSSLTKSODLAYVIFTSGTGDPKIMI 653
DB 11700 VKUHANV-----RFRISDALVESGPPTEELSTRPTAQSLAYVFTSGTGPVKGMV 11754
QY 654 EHRAFSSCALFKGASLGINSIDRALQ-----FGTHAFGACLEIMTITLINGG---CV 702
DB 11755 EHRGIT-----RLVKNSNVVAKOPAAAAIAHLSNIAFDASWEIYAPLLNGGIVVCI 11806
QY 703 CIPSDDRMNSIPSFNRYNVMNMATPSYMG---TFSPEDVPLGLATLVLVGQMSSSVN 759
DB 11807 DYTITID-IKALEAVFKQHIRGAMLPPALLKQCLVASPTMISSLEILFAAGDRLSSQ- 11864
QY 760 AIWAPKL--QLLNGYGQSSSSICFASN-STEPNNMGAVG-----AHSWIDPDNDIN 810
DB 11865 AILARAVSGVYNAVGPENTVLSTIHNIGENAEFNGVPIGNAYNSGAFVMDQN--Q 11922
QY 811 RLVPIGAVGLVIESPGIARDYIVPPPEKSPFTDIPSMYPAINTF-----PDG--AKLYR 864
DB 11923 QGLSAGVIGELVYTGDLARGY-----TD--SKLRVDRFTIYITLDGNRVAYR 11968
QY 865 TGLDARY-ASDGSIVCLGRIDSQVKRGORVELGALETHL--ROOMPDLITVVEATKRSQ 922
DB 11969 TGDVRHRPKDQIEFFGMDQDQIKIRGRIEPAEVEQALARDPAISDSAVITQLTDEEE 12028
QY 923 -----SANSTSLAFLIGSSYFGNRPNSDAHILDHATKA-----INIKLEOVL 965
DB 12029 PELVAFESLKGANGTNGV-----NGVSDQEKIDGDEQHALLMENKIRHNLOALL 12078
QY 966 PRHSIPFYICMLELPRTATGKIDRRRLRMKGDIIDKOTQGAIVQOAPAPVPFADTAA 1025
DB 12079 PTMIPSRIIHVDPVANGKIDRNELAVR-----AQATPRTSSVSYVAP 12125
QY 1026 K--LHSIWQSLGIDPATVNGAT--FFELGNSITAIKM-VNMARSVGMDLKVSNIQH 1080
DB 12126 RNDIETIICKEFA-DILSVRVGITDNFFDLGGHSLIATKLAARSLRLDTRVSRVDFDT 12184
QY 1081 PTLAGISAVKVGDPVLTPIKSTHEGPEQYSQGRMLFQLDQVGLSWLIPYAVRMR 1140
DB 12185 PVVQOLAASIQOOSTHEALPALSHPGVPQOQPAQGLWFLDFRLNNAWYIMPFVGLR 12244
QY 1141 GPNVDALRALAALBQRHETLTTFEDQGVGVQIVHEKLSBEMKVIDLCGSDLDPEFV 1200
DB 12245 GPLRVDALQALRALBQRHETLTTFEDQGVGVQIVHSPMRDICVVDISGANED-LAK 12303
QY 1201 LNEQOTTPNLSSEAGWRATLLRLGEDDHLITIVMHIIISDGWSIDVLRDLNQLYSAAL 1260
DB 12304 LKEEQAPNLSPEVAMRVALLFKAGENHHILSVMHIIISDGWSIDVIFQOELAQFYSAV 12363
QY 1261 KDSKDPALSALTPIQOYSDFAKWQKD--OFIEQEKOLNWKKOLKSDSPAKIPTDFARPA 1318
DB 12364 R-CHDPLSQVKPLPIHYRQFVAVNQKQVAVHESQLOVWIEGLADSTFAEILSDNRPE 12422
QY 1319 LLSGDAGCVHVTIDGELYSLRAFACNEHNTTSFVLLAAAFRAAHYRLTAVEDAVICTPTA 1378
DB 12423 VLSGEAGTVPVIEDEVEYKLSLFCRNHQVTSFVLLAAAFRAVHAYRLTGAEDATICTPTA 12482
QY 1379 NNRPELEDILIGCVNTQCMRINIDHHDHTEGTLINOVKATTAFAFENEDIPFERVVSALQ 1438
DB 12483 NNRPELEDILIGCVNTQCMRILEEHNDLNSVVRVRSTAAAFENQDVPFERVLSALQ 12542
QY 1439 PGSRDLSSTPLAOLIFAVHSOKDLGRKFGOGLSVSPVPSKAYTRFDEMEFHLFOETDSLKG 1498
DB 12543 PGSRDASRNPVLQMFVHSQRNLGKLQLEGEGEPTPYATTFRFVFEHLFQODKGLAG 12602
QY 1499 SVNFADLFKMEIVENVVRVFFETLRNGLOSSRTPVSIPLDTGIVTLEKLDVLNVKRV- 1557
DB 12603 NVVFAADLFPAATIRSVVEFHEILRGLDQPDIAISTMPLVDGLAALNSRNLPAVEDIE 12662
QY 1558 -DYPRSSSLADVQTVQVSAYPDSLAVVDSRCRTIYTELDRQSDILAGLWRRRSPARTLV 1616
DB 12663 PDFATEASVVDVFQTVQVNPALAVTDTSTKLTIAELDQSDHVAWLSKQKLPASIV 12722
QY 1617 AVFAPRSCETIVAFPGVVKANLAYLPDVRPSARVODILSLSGPTVILIGHDTAPDI 1676
DB 12723 VVLAPRSCETIVACIGLILKANLAYLPDMSNVPARRQAILSEIPGEKVFLLGAGVPIPDN 12782
QY 1677 EVTNVEFVRIRDALNDSNADGFEVIEHDSKPSATSLAYVLYTSGSTGRPKGVMIHRVI 1736
DB 12783 KTADVRVFTSIDIVASKTDKSYS-----PGTRPSASSLAYVIFTSGSTGRPKGVMEHRV 12838
QY 1737 IRVTSGCIPNYPSETRMAHMAITAFDGSYEIYSALLFQRTLVQVDYMTLDRALKDV 1796
DB 12839 ISLVKQNA-SRIPQSLRMAHVSNLAFDASVMEFTLLNGTGLFCISYFTVLDSKLSAA 12897
QY 1797 FFEHVNAAASHVTSSSQDVLVRPRLSRTLMEFFLVVTDSTAPDALDA--QGLYQG---- 1851
DB 12898 FSDHRIN-----ITLLPALLKQCL-----ADAPSVLUSSLESYIGDRL 12937
QY 1852 -----VQCYNGYPTENGVMSTIYIPIDTESPIINGVPIIGRALN-NSGAVVVD 1897
DB 12938 DGADATKVKDLVKGKAYNAVGPSTENSVMSTIYIEH-ETFANGVPIGTSGLPKSKAYIMD 12996
QY 1898 PEQOLVIGVGMELVYTGDLARGYSDKALDENRFRVHITVNDQTVKAYRTGDRVRYRIGD 1957
DB 12997 QDQOLVPAGVGMELVAGDLARGYTDPSLNTGRFTHITIDGKQVAYRTGDRVRYRPRD 13056
QY 1958 GLTEFFGRMDTQKIRGNRIESAEIAALLRDSVRDAVVVLOQNEQAPFELGFFVADV 2017
DB 13057 YQIEFFGRDQOQIKIRGRIEPAEVEQALLSSINDAVVVSQNK-EGLEMGVYITQA 13115
QY 2018 DHSENDKGSANVEGQDHFESGMYSDIGEIDPSTIGSGKWTSMYDQSIDFDBMHE 2077
DB 13116 AQSV-DKEEASNKVQEWAEHFDSTAYANIGGIDRDALQDQFSLWTSYDGLSLIPREMQE 13174
QY 2078 WLGETTTLHDNBSLGNVLEIGTSGMILFNLD--SPLESYVGLSPSRSAAFVKNKATES 2135
DB 13175 WLMDTRSLLDNPPGKVLGIGTGMVFNLGKVEGLQSYAGLEPSRSTAVWNKALET 13234
QY 2136 IPSLAKAKYQVQVQATDIGOVDLHPDLVVLNSVIQYFPSPSEYLAETADTLHLPNVQRI 2195
DB 13235 FPSLAGSARVHVQAEEDISSIDGLRSDLVINSVAQYFPSPREYLAETLANLRPGVKRI 13294
QY 2196 FFGDVSQATNEHFLAARAIHTLTKGNATKDDVKQMAELMBEELVBEPAFTSLKDRF 2255
DB 13295 FFGDMRTYATNKDFLARAVHTLGSNASKAMVQQVAKLEDEDEELVDPAPFTSLSDQF 13354
QY 2256 PGLVEHVEILPKNMEAVNELSAYRAAVHVHVRGSL---GDE---LVLPEXDDWDIDFQ 2307
DB 13355 PDEIKHVEILPKRMAATNELSSYRYAAVTHVGHPNGMPNGDEDEKQNAVKDINPKAWVDFA 13414
QY 2308 ANGLNOKSLGDLKSSDAA--IMAVSKIPETIFAFERQVVASLNSNID-----EWOL 2357
DB 13415 GTRMDQALLQLLQDRQDGDVVAVSNIPYKTIHERHLSQSLDDDDGTSVAVDGTAW-I 13473
QY 2358 STTRSAEGDSSLSVDPDIFRIAGEAGFRVEVSSARQMSQNGALDAVPH-----HCCSQG 2411
DB 13474 SRTQSRAKECPALSVADLIEIGKIGIGFEVSEASWARQHSQRGLDAVHFRPEPRH---SG 13530
QY 2412 RTLNVFPTDHLHKGSDLLTNRPQRQRRIAEVRELRSLPSYMIPSNIVVLDKMP 2471


```
Db 164 YVMPARITLIDQMLNANGKVDKDLARRAQTVSK--ABKLPASRVAPRNEVEVLCDEF 221
QY 1021 ADTAALKHSIWQSLGIDPATVNVGATFEELGNSITAIKM-VNMARSVMCDLKVSNIYQ 1079
Db 222 SDV-----LGEV---GVADSFDFLGGHSLMATKLAARISRRNARNARVKEVDF 267
QY 1080 HPTLAGISAVVGDPLSYTLIPKSTHEGPEVQSYSGRLWFLDQDVLGSLWILIPYAVRM 1139
Db 268 QPVADLAIVIQSGKPHNPILTAPVSGPVEQFAGRLWFLDQINLGASVLMPLAVRL 327
QY 1140 RGPVNVADLRALALEORHEHILRTTFEDQDGVGVQIVHEKSEEMKVITDLGCSLDLP-- 1197
Db 328 RGPLHEALTITAILALEORHEHILRTTFEDQDGVGVQIVHEKSEEMKVITDLGCSLDLP-- 1197
QY 1198 FEVLNQEQTPNLSSEAGWRATLRLGDGDLTLITVMHHIISDGWSIDVLRDLNQLYS 1257
Db 387 LQPLKQEQSSFDLASEGWRVSLRLRIANDHVLISVMHHIISDGWSIDVLRDLNQLYS 446
QY 1258 AALKDSKOPLSALTPLPIQYDFAKWK--DQFIEQKOLNWKYKKOLKDSKPAKPTDFA 1315
Db 447 AALR-GYDQSQSLPLPIQYDFSVNKKQAEQVAEHERQLEYWTTTLADDSPAELTDL 505
QY 1316 RPALLSGDAGCVHTIDGELYSLRAFCHENHTTSFVLLAFRAAHYRLTAVEDAVIGT 1375
Db 506 RPTVLSGNAGVYQLADIGSLYKLAFCRAYQTTTFSAVLLAFAFRATHYRLTGAEDATIGT 565
QY 1376 PIANRRNPELEIDIGCFVNTQCMRINIDHHDTFGTLINOVKATTTAAAFENEDIPPRVYS 1435
Db 566 PIANRRNPELENLIGFVNTQCMRIIVDRDDTFEILMQQVSTTTAAFAQDVPPFERIVS 625
QY 1436 ALOPGSRDLSSTPLAQILPAVHSQKDLGRFKQGLSVSPVSKAYTRFDMERHFLFOETDS 1495
Db 626 ALLPGSRDLSSTPLAQILPAVHSQKDLGRFKQGLSVSPVSKAYTRFDMERHFLFOETDS 1495
QY 1496 LKGSYNFADLEKMEIVENVVRVFFELLNGLOSSRTPVSIILPTDGIIVLEKLDVLNVK 1555
Db 686 LGGTVLFAFDLFEPPDIRGMVTVFQEVHLHVLDQPHPTITSPLANGLAGLCNMGLLDIE 745
QY 1556 HVDYPRESSLADVFQTSAYPDSLAVDSSCLTYTDELDRQSDILAGWLRRRSPAEATL 1615
Db 746 RSDYPRESSLADVFQTSAYPDSLAVDSSCLTYTDELDRQSDILAGWLRRRSPAEATL 1615
QY 1616 VAVFAPRSCETIVAFPGVLKANKAYLPDVRSPARKVDILSGSLGPTIVLIGHDTAPPD 1675
Db 806 VGYLAPRSCQTVAFGLILKANLAYLPDWNVPAARIEAILSAVEGHKLVLGSDVAAK 865
QY 1676 IEVTVNVEFVIRDALNDSNADGEVIE---HDSKPSATSLAYVLYTSGTGPKGVMI 1732
Db 866 VRQDVEVVRVDDTL-----GYDILDDYANSAPRSPSTSLAYVMTSGTGPKGVMI 919
QY 1733 HRVIRTV-TSGCIPNYPSETMAHMTATIAFGASYEISALLFGRTLVCVDYMTTLDR 1791
Db 920 HRVIRTV-TSGCIPNYPSETMAHMTATIAFGASYEISALLFGRTLVCVDYMTTLDR 1791
QY 1792 ALKDVFFRHHVNAASHVTSQDQVPLVRPRRLSRTLMLFFLVVTDSTAPDALDAQLYOG 1851
Db 980 ALETVMQQRVAMLTPLALLKCLADIPSLVG-SLDVLFNVGDRPDRSDALAAQAL--- 1035
QY 1852 VOC--YMGVGPTEGV 1865
Db 1036 IRGVVYNAYGPTENGI 1051
```

RESULT 5

Q9L8H4

ID Q9L8H4 PRELIMINARY; PRT: 4247 AA.

AC Q9L8H4;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Actinomycin synthetase III.

GN ACNC.

```
OS Streptomyces chrysomallus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1895;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 11523;
RX MEDLINE=20243879; PubMed=10780924;
RA Schaecker F., Pfennig F., Grammel N., Keller U.;
RT "Construction and in vitro analysis of a new bi-modular polypeptide
synthetase for synthesis of N-methylated acyl peptides.";
RL Chem. Biol. 7:287-297(2000).
DR EMBL; AF204401; AAF42473.1; -
DR HSSP; P14687; IAMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR001601; Methyltransf.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR000051; SAM_bind.
DR InterPro; IPR000379; Ser_estrs_site.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00501; AMP-binding; 3.
DR Pfam; PF00668; Condensation; 3.
DR Pfam; PF00550; pp-binding; 3.
DR Pfam; PF00975; Thioesterase; 1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00075; ACP_DOMAIN; 3.
DR PROSITE; PS00455; AMP_BINDING; 3.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_3.
KW Phosphopantetheine.
SQ SEQUENCE 4247 AA; 462293 MW; 580A7F41522A0BC5 CRC64;
```

Query Match 14.8%; Score 2391; DB 2; Length 4247;
Best Local Similarity 26.7%; Pred. No. 2.3e-138;
Matches 888; Conservative 488; Mismatches 1282; Indels 664; Gaps 117;

```
QY 41 LSSRIEAKPCTPFOLDIMDCNALDKQSAIGHA---VVDPTDIDISRALAWKEIVNQ 97
Db 1 MTSSGLLEILPLSPMQEGLLFHSRYEQDGDVAVQVHVLDLGDLGRLREAVAGLVOR 60
QY 98 TPALRAFATSDSGKTSQVILKDSFVSNMCMSSSSP---DEVVRDEAAAAAAGPRCN- 153
Db 61 HPLNRAAFQVDSGRVQLISR-HVDLPWEEFDLSSMPAPEAEALSRIAAKCHGRFDP 119
QY 154 -----RFVLEDMQTKKQOLVMTFSHALVDVTFQORVLSRVFAAYKHKEDTHRPE-TPE 206
Db 120 AEPPLRFTLVRTAE-EHRLILTHILLDQWSTPLVRELTAADGPAALPRVTPY 178
QY 207 SSDATDTSQSVSVSMSCEDNAVSAHFQWTHLNDLNASVFFHLSDHLMVP-NPTTTA- 264
Db 179 R-----QYLGWLAQQDRPAAEAAMREALGDLEQPT-----LVAPVDPARAAL 220
QY 265 -EHRITFPLSOK-----ALSNSAICRTALSILLSRYTHSDEALFGAVTQSPLP 311
Db 221 MPERITELAEETAALVEWARHGVTLNTVLOAAGLVLSRRTGTHHVVFGVAAGRDP 280
QY 312 FDKHYLAD-----GTQYTVAPLRV---HCOSNL-----RASDVMDAISYDDRLGHLAPFG 359
Db 281 ----QLAGVESVMVGLLITWVPRVRLDHAGSLLEAVRRLQDEQSRLTSHH----HLGIAR 332
QY 360 LRDIRVTGNGSAAQDFQ-----TVLLVTGDSHVNNINGFLOQITESSHF-- 405
Db 333 IQOLTGMSELFDTSLVFENYPMQDPAELPATGLRITPD-----LGRGDATHYPL 382
QY 406 --MPCNNRALLHCQMESSGALLVAYVDHNVDSLQTTRELQOFGHLIKCLQSPDLDS-- 461
Db 383 TLIAAPGRRLYLRLD-----YRDLFRAEACALVER---LIRV----LDLAVA 424
QY 462 ----SMAEVLNMEYDRAET---ESWNSOPLEYQDVTLIHHEMLKA-VSHSPTKTAQAWD 513
Db 425 DAERPLGRVDLLTAERRELLHAGNGAGMPLAASL---PELFRARVAAPDAVAVEYGE 481
QY 514 GOWTYSELONVSSRLAVHKSGLGRAQQAIIPIVFEKSKWVIASMLAVLKSNAFTLIDP 573
```

Db 482 EALRYELDERANLAHLAAHGV-GPERIVALALPRSDVLVAVLAVLAKAGAAAYPLDP 540
QY 574 NDPARTAQVOTRATVALTSKLRHETVOKLGR--CVVDDDELQSVS---ASDDFS 627
Db 541 EYPANRLAHWTDQPTVLTT---TETEAKLPDRHPCILLDDPAVLADLSGRPAHDPV 597
QY 628 SLTKSDLAYVFTSGTGPDKGIMI-----EHRAFSSCALKGASIGNSDTR 676
Db 598 ELHPDHP-AVIVTSGTGPVPGVMPAGGLNLLQWHRA-----VGDEPGR 645
QY 677 ALQGFTHAFACALLEIMTTLINGCCVCPISDDDMNS--IPSEINRYNWNWMAATPSYM 734
Db 646 TAOFTAISFVDSQAEVLSSAFAGKTLVPEEYVRRDAAARFAGLDDQVDDELFAPNLVLE 705
QY 735 TSPEDV-----PGLATLVGE--OMSSSVNAIW--APKLOLLNGYGOSSESSICFAS 784
Db 706 ALAEAAVETGRTLPOLRTVAQAGEALTLSRTVAFHRSAPGRRLHNNHYGTE-THVVTAH 764
QY 785 NMSTEPNN-----MGRV-GAHSWIDPNDINRLVPIGAVGELVIESPGIARDYIVPPP 837
Db 765 ALGDDPEDWRLPAPIGRPIDNTHAY--RTRAVRLVBPVGVGELYIAGAGLARGYLGRPA 822
QY 838 PEKSPFTDIPSWYPANTFPGAKLYKTGLARYASDGSIVCLGRIDSQVKIRGORVELG 897
Db 823 LTAERFVAD-----PYGLEP-GGRMYRTGDLVRNPDGELEFCGRADHGVKVRGRIEP 876
QY 898 AIETHLRQOMPDDLTIVVEATKRSQSANSLSLAFILGSSYFGNRPDSAHILDHDKATKAI 957
Db 877 EIEKVLTDH-PDIAQAAR--VTRPHRPGDRLVAYVVGREAL--RP-----917
QY 958 NIKLEQV-----LPHSPISFYICMLPRATGKIDRRRLRIMGKIDLKQTOGAIV 1010
Db 918 -----EQVREFTRERLPHMVPAAVQLERLPTPNGKLDRAAL-----956
QY 1011 QQAPAPVFPADTAALKHS-----IWQSLGIDPATVNVGATFELGGSNTIAKMW 1062
Db 957 ---PEPDAFALAGCREARTPOEIVCDLFAQVIGL--PWVGVDDEFELGHHSLLATRLI 1011
QY 1063 NMAVS--GMDLYKSNIYQHTPLAGISAVVKGDPVLSYTLIPKSTHEGVEQSYQGRLWFL 1121
Db 1012 ARIRAFSVELGLRTLFEARTAAVAHAHLDLAGPARTALTKHQLPDAVPLSFAQRLWFL 1071
QY 1122 DQLDVGLWTLIPYVNRMPGVNDALRALALEQRHETLRTTFDQDGVGVQIVHEKL 1181
Db 1072 HKMEGPSATNIPILALRLTGELNLEALRALEDVOTGRHESLRTVFPEVDGTPYQRLDTA 1131
QY 1182 SEEMKV1DLCGSDLPPEVLNQETTPFNLSSEAGWRATLLRGEDDHILTIVMHHITSD 1241
Db 1132 TIEL--VRTATTAEELTDHLRAARHFPDLAGEPPLTELFSAREHVLVLLVHHIAGD 1189
QY 1242 GWSIDVLRDLNOLYSAALKSDPLSALTPPLTIQYSDFAKWQDQFIEQ-----KQ 1294
Db 1190 GWSIGPLASLARAY--AARAGRAPQWA--PLPVQVADYTLWQNELLDGQNDPSLFPATQ 1246
QY 1295 LNTWKQOLKD--SSPAKIPDFAFPALLSGDAGCVHVTIDGELYQSRAFNEHNTSFVV 1353
Db 1247 VAYWETLAGLPQITLTPDRPFAVMVTRYGDYLTVDIDPELHRLTELARGSGASLEW 1306
QY 1354 LLAFAAAHVRLTAVEDAVIGTPIANRNREPELIEDIGCFVNTQCMRINIDHHDTFGLIN 1413
Db 1307 LQAGLAALIKRAGDDIPGSIAGRTDQALDOLGFFVNTLVLRITDITGDTFTQLLT 1366
QY 1414 QVAKTTAAFENEDIPFERVVSALQPGSRDLSSTLAQLIFAVHSQKDLGRKFQGLSV 1473
Db 1367 RVRETSAAVAHQDVPFYLVEVLNP--TRTLAHPPLFQIMLALQNAPE-GTFQLPGLTVD 1424
QY 1474 PVSKEYT-RFDMEFHLQ-----ETDSLKGSVNEADELFKMETVENVVRVFEILNRG 1526
Db 1425 VAPGRTGTAKFDFFLSLAERRGADGEPQGITGAVEYSSDIYDAPTQALFNRIHLDTA 1484
QY 1527 LQSSRTPVSLPLTDGTVLTKELDLVNRKHVDYPRE---SSLADVFQTOVSAYPDSLAVY 1583

Db 1485 TAOPEOPLSRIDL-----LTAQEHQLDLDTWLDTAVEVGPDLPLPARFARQAATPQAVALI 1540
QY 1584 DSSCRITYELDRQSDITAGLWLRSSMPAETLVAVFAPRSCETIVAFGVYKANLAYLPL 1643
Db 1541 AGDLSLYAELDARANRLAHALLREGAGPDRBLVALALPRTAELVALLVLTGCAAYLPL 1600
QY 1644 DVRSAPARVODIILSGSGPTIVLIG-----HDTAPPDIEVTNVFVRIRDALND-SNADG 1697
Db 1601 DPDHPAARLSHVL-GDARPALLLTDRTEQHLPADADTRRLALDLSAEVALLADCPOTDP 1659
QY 1698 FEVLEHDSKPSATSLAYVLYTSSGCRPKGVMLHVRVIRTYS--GCIPNYPSEPRMA 1755
Db 1660 AE-----EGVTPAPCSAAIVITYTSGTGRPKGVVPHSALNVFVTAMRRQAPLQOE-RLL 1714
QY 1756 HMTIAFDGASYEILYALLFRTLVCDVYMTTLDARALKDVFREHNAASHVTS-----1810
Db 1715 AVTTVAFDIAALELYHPLLSGAAVVLAPKAEPQPSAVLDLIARHGVTTVQGTPLMQLL 1774
QY 1811 -----SSQDVLPRVPRRLSRTLMMFFFLVLTDDSTAPDADLAQGLYQVOCY 1855
Db 1775 VGHDAEALRGLRMLVGGEALPLSLAEALR-----ALTDOLV-----1810
QY 1856 NGYPTENGVMSTIYPIDSTESFINGVPIGRALNNSGAYVVDPEOQLVGVGMGELVVTG 1915
Db 1811 NLGYPTETIWTAAEL---AGGTGAAPIGRPIANTRVYVLDLQLOPAPVGVWGLYIAG 1867
QY 1916 DGLARYSDK-ALDENRFVHITVN-DQTVKAYRTGDRVYRIGDGLIEFFGRMDTQFKIR 1973
Db 1868 AGLARYDLRPAALTAERFPADPYCLEPGGRMYRTGDLVRN-PDGELEFFVGRADHVKVR 1926
QY 1974 GNRIESAEIPALLRSSVRDAAVVOONEDQAPETLGVFVADHDSHNSKGSANOVG 2033
Db 1927 GFRIEPEIEKVLTDHPDIAQAAVVRREDQGDARLVAYVVTGGSADARQEQEQDQDE 1986
QY 2034 WQDHESGMYSDIGEIDPSTIGSGDKWTSMYDSQIDDFDEMHEWLGETT---RTLHNR 2090
Db 1987 WQDLYDS-VYTAASE---TAFGENFASWNSYDGLPIPLDOMREWRDTTVERINGLPRR 2042
QY 2091 SLGNVLEIGTSGMILFNLSRLSRLSYVGLPEPSAAAFVANKATESIPSLAGKAKVQVGT 2150
Db 2043 ---VLEIGVGTGLLSRLAPHCCEYWGTFDPTVIADLRGHVEADPELAARVQLRTQA 2098
QY 2151 TDIGQVDDHLPDLVNLNSVIQYFSPSEYLAETDLIHL-LPNVQRIFFGVDVRSQATNEHF 2209
Db 2099 HFDQDLPHGHFDVWNLNSVQYFPNAGYLEQVLDHALRILAPGTVTFIGDIRNPLRLTF 2158
QY 2210 LAARATHTLGKNATKDD--VRQKMAELEDMEELLVPEAFFTSLKDRFPGLVEHVELTPK 2267
Db 2159 --TTAVHTARATPDADHAARAVRAVEHALVEKELLIDPEYFTALGHVHPDLA-GVDIRLK 2215
QY 2268 NMEAVNELSAYRAAVVHVRGSLGDELVLPEVKDDWIDFQANQLNOKSLGLLKSDDAAI 2327
Db 2216 PGTVENELTRYDATLHKAGATHPLTGPPER-PWSGHADPDALAEALARQLDSERPAR 2274
QY 2328 MAYSKIPFETATERQVVASLNSNIDENWQLSTRSSAEGSSLSVDP---IFRIAGEAGF 2384
Db 2275 LRVGTGPNPLARELALQHALDT-----ATAPISPTADAAQGPQDLDALHRLGRQHY 2329
QY 2385 RVEVSSAROWSON--GALDAVFHCCS-QGRTLVNPFTDHLRGSDDLTRNRLQRLNRR 2441
Db 2330 ----WTAAHTNTHRPDAVDLVVALSDLDGAT-----PTGTYSAGCTSTPATPLSSLTNP 2381
QY 2442 IA-----IEVRERLSLLPSYMIPIINSIVLDMKPLNANGVKDRKELSRRAKVVPKQ 2493
Db 2382 AAGRGTSALLTTUREHARTHLPDYMQPSALVPLDRLPLTANGKLDL-----2427
QY 2494 TAAPLTPFPISEVEVILCEEATEVFGMKVDITDHFHNLGHSLLATKLISRIDQRLKVI 2553
Db 2428 AALPAPDFTLA-----2438
QY 2554 TVKDFDHPFVADLASVIRQGLQOPVSDGQDRSAHMAPRTETAILCDEFAKVLGF 2613
Db 2439 -----GTGRE-----PRTPOQIVCDLFTQVLGL 2462


```
QY 57 LMDICNALDKOSAIGHAVYDVTIDISRF-----AL--AMKEIVNOTPALRAFAFTSDS 110
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2177 LDQLQPN-----SLYNIPLALRLAGLEVALEQSLLEEITYRHEALRNFITID- 2226
QY 111 GKTSQVILKDSFVFWMC-----WSSSSSDEVVYRDEA-----AAAASGPRCNR 155
Db 2227 GKPTQII---QTVRNQWLSVIELQHLPRLEQETSQAELAITAQIOPFDLASGLLRATLV 2283
QY 156 VLLEDMQTKKCOLWTFSHALVDVTFOORV-----LSRVFAAYKHEKDHTRPETPSSDAT 211
Db 2284 VLSE---TEHILLI---CHHIVISDEWSGVFTQELATSYNAYAQOSANTAPLP----- 2332
QY 212 DTFDSOSVSVMSCEDNAV-----SATHFWQTHLNDLNASVPHLSDBHLMVEN 259
Db 2333 -----IQYADFATWQRLVGEVLQSQLSYWKQQLKADAPAL-----LSLPTDRPR 2377
QY 260 PTTT-----AHRITFPLSOKALSNS-----AICRTALS-----ILLSRYTHDEALFGAV 305
Db 2378 PAVQSLGAYQEFALSLELTQGLMQLSQOQGVTFLEMTLLAGFDTLTRYTGOSDILVGP 2437
QY 306 TEQSLPFDKHYLDAGTYQTVAPLRVHCOSNLRASDVMDAISYDDRLGHL----- 355
Db 2438 -----TANRHHSEIEGLGFFVNTLVWRTDLSADPFSLLGVRREMAISAYAH 2486
QY 356 ---APF-----GLURDINTGDNGSAACDFQTVLLVTDGSHVNNNGINGFLQOITESSHFMC 408
Db 2487 QDLPEMELVEALQPNRL-----SHTPLFOVAFVQNPAPNSQVELTG-----LTVSPLVRES 2538
QY 409 NNRAILLHCOMESSGALLVAYYDHN--VIDSLQTRLLQOQGHLLK-CLQSPDLUSSMAE 465
Db 2539 TTAQFDLTLAMENTGLGVWYENYDIDFDSSTIERMAHSHFVTLLEGIVANPFERISQLP 2598
QY 466 VNLMTXDRAETESNNSOPLEY-QDTLIHHEMLKAVSHSPTKTAQWGDWNTYSSELDNV 524
Db 2599 LITAVEQOOLLIE-WNDTQVDVPQDKCHOLFEEQCLRTPDAVAVFENQOQTYHELNCR 2657
QY 525 SSR LAVHKSGLRAQQAIIPIYFBEKSWKVIASLAVLKSNGNAFTLIDNPDPARTAQVY 584
Db 2658 ANQLAHYQLSLGVGA-DVLVGLCVERSLEMIVGLLAILKAGGAYVPLDPDPYQERLNLIL 2716
QY 585 TOTRATVALTKLHRETQKLVGR----CVVVDDELQSVASDDFSSLTKSQDLAYVI 639
Db 2717 EDAQVSVLL---IQOHLVEKLOQHOAHVCLDSDGKEKTAQNSNPLNATIPES-NLAYVI 2772
QY 640 FTSGSTGDPKIMIEHRAFSKALKFGAS----LGINSIDTRALQFGTHAFGACILLEMTPL 696
Db 2773 YTSGSTGPKGVLVNH---SHVVRLEAATDSWYNFNSQDVWTFMHSYAFDFSVWEVWGAL 2829
QY 697 INGGCVCI-----PSDDDRMNSIPSFINRYNWNMMATPSYMGTFSPEDVP----- 742
Db 2830 LYGGRLVVGYLVTRSP-----KSFYELLQCEKVTILNQTPSAFRQLIPAEQSIATVGD 2883
QY 743 -GLATLVLVGEQSSSVNAIWAPK-----LQLLNGYGOSESS-STCF-----ASNSTEPN 791
Db 2884 LNLRLVIFGGETLEINSLOPWDRHGDSQPLVNNYGITETTVHTVYRPLSKDALHGKAS 2943
QY 792 NMGRAGV-AHSWVIDPNINRVLPIGAVGELVIESPGIARDYIVPPPEKSPFFTDIPSW 850
Db 2944 VIGRPIGLQVVLDEH--LQVPPIGVAGEMVVGAGVTRGYNLNRAELTAQRFI----- 2995
QY 851 YPANTFPDGAK--LYRTGDLARYASDGSIVCLGRIDSQVKIRGQVRELGAITHL----- 903
Db 2996 --SNPFNGNSEOLLYKSGDLARYLPNGELEYLGRIDNQVKIRGFRELGEIEAALSQLE 3053
QY 904 -----ROQMPDDLTV-----VEATKRSQSANS----- 926
Db 3054 VREVVVVARSQDPQDNKRLVAYVVPQOKNLESSKPEKLTSDKVELWPSIAEYVYVDDLLY 3113
QY 927 ----- 926
Db 3114 YAMTNDHRRNDSYKVAINQLVKDKIVVEIGTGKDAILSRFCVOGGANKIYAIERNBETCR 3173
```

```
QY 927 -----TSLIAFLIGSS----- 937
Db 3174 QARACIEKGLADKTVIHGDATLVDIPELADVCVSEIVGAIGGSBGAIVINNRRFLK 3233
QY 938 -----YFGRN----- 942
Db 3234 PDGLMIPERSITKMAAVTLPDEILHNPKFAEIPAYYTKKIFEEVGYFPDLRVCIKFSQA 3293
QY 943 -----PSDAH-----ILDH-- 951
Db 3294 NVLSTVDVLEDLNFNEYISPEFSHEVQLEIHKNGRMNGFLVWLNLTHTIAGEEDILLEHEY 3353
QY 952 -----DATKAI----- 957
Db 3354 CWLPVPLPFEFIEVEYGDVIAVCVSRCLCNENLNPYAIKGRLLKNGENINFEYVSY 3413
QY 958 -----NIK-----LEQVLPHRHSIPSYFICMLPLPRT 983
Db 3414 HYKNNFRQTRFYORLFADYGNVKNYYPDNLARYLRANLKKLLPDYMIPISEVFNLENDLPT 3473
QY 984 ATGKIDRRRLRMGKIDLDKOTQGAIVCOQAPAPIPVFADTAAK--LHSTWQVSLGIDPAT 1041
Db 3474 PNGKIDHLAL-----PKPELTVLSEADYVMP---NTEAKIITAAWQKILQIE--K 3519
QY 1042 VNVGATFFELGNGSITAIKMN--MARSGMDLKVSNYIHOHTLAGISAVV-----KCDPL 1095
Db 3520 VGLDNFPELGGHSLLATQVISRLQEAFGISLPLRYIFESPTVAQLSEVILSQTGSGSL 3579
QY 1096 SYTLTPKSTHEGPVQSYSGQRLWFLDQDVGSLWTLPIYAVNMRGPNVNDALRRALAAL 1155
Db 3580 TVPAIAPVDRDRDIPLSWAQERLWFVNOLEGESSAYTIDFTMRLVGNLAVKALEQAFAAI 3639
QY 1156 EORHETRTFFDQDQGVQIVHEKLEEMKVIDLCSGLDLP-----EVLNQBOTPPFNL 1211
Db 3640 GORHEPLRTOFKIKDNKPQVATAPSWTITLPPVDLQNLQ-DPTKQVEELATAEACPFDL 3698
QY 1212 SSEAGWRAPTLRLGEDDHITIVMHIIISDGSVIDLRRDLNQLYSAALKSKDPLSALT 1271
Db 3699 ANGSVLRVLQWQVAPDEYVLLLAIIHAADGSMGILIDELSAYYRSFSTGSSVELPELS 3758
QY 1272 PLPIQYSDFAKQKQOFIEQ--EKQLNYWKQKLDSSP-AKTPDFARALLSGDAGCVH 1328
Db 3759 ---IOYADFAVWQORWLTNQVLEQCQSYWKQLTGAPLLELPTDRPRAIOTFRGTEQ 3815
QY 1329 VTIDGELYQSLRAFACNEHTTSFVLLAAFRAAHYRLTAVEDAVICTPIANRNPBELEI 1388
Db 3816 LQDLSOLTQOLKLSQKSGSTLEMTLLAGFVVLLSRYSGQTDLVGSPATANRNOQIEPL 3875
QY 1389 IGCFTVNTQCMRINIDHDDTFTGLINOVKATTTAAAFENEDIPFVVVSAJOPGSRDLSSTP 1448
Db 3876 IGFVNTLALRFDLSPEPSFEALLAQVRKVTONAYEHQDLPFEMLEELQL-ERNLDRNP 3934
QY 1449 LAQLFAVHUSQKDLGRFKQGLSVVPVPSKA-YTRFDMEFHLFOETDLSLKGSVNFADEL 1507
Db 3935 LTQVVFALQN-APTAPWDPFGRVQDMPSGLDOSVRVDLELYLWDVPEGLGGFCYSNRDLF 3993
QY 1508 KMETVENVVRVFEILLNGLQSRTPVSTILP-LTDGIVTLEKLDVL---NVKHVDYVPRES 1563
Db 3994 DGTITARMQNFVTLGALVNDNSQQSVALLPLLTGEG---DRHQLLKENDTQTDOPYHK 4049
QY 1564 SIADVFQTVQVAPYPSLAVVDSRCLTYTTELDQRQSDILAGWLRRRSMPEATLVAVFAPRS 1623
Db 4050 C1HOLVEEKAQETPDATIAVVCNQQLTYAQLNRRANQLAHLYRSLGLETEQLVGJCVRS 4109
QY 1624 CETIVAFFGVKANLAYPLDVRSPSARVQDILSG-----LSGPIV--LIGHDTAPPD 1675
Db 4110 PLMLVGIILGILKAGGAYVPLDPEYPOERLSLMELEDSQVSLLLTQMTLIEKLPQHPAKPV 4169
QY 1676 IEWTNVFVRIRDALNDSNADGFVIEHDSKPSATSLAVLYTSGTGPRKGMVMEHRV 1735
Db 4170 LD-----EIQOQIAQNSQENLTGI-----VTAANLANVYTSSTGKPKGVVMEHOG 4216
QY 1736 IIRTVTSGCIP-NYPSETRMAHMATIAFDGASYEYISALLFGTLVCVDYMTTLDA---- 1790
```

```
Db 4217 LCNLAQAQIAQFGLHSDSRILOFASLDFACISEILMSLGSQGLT-----YLGSKDSLMPG 4272
QY 1791 ----RALKQVFFREHYNAAASHVTSQSDQPLVRPRLSRTLMFFFLVTDVSTAPDALDAQ 1846
Db 4273 TPLIERLDY-----ATHIILPSS--ALAVLHKVELTTLTIIVAGEACAVELIKQW 4323
QY 1847 GLYQGVQCYNGKGTGPTNGYMSIYIPIDSTESFNGVPIGRALNNSGAYVDPPEQQVLVGIG 1906
Db 4324 S--AGRNFFNAGPTGEGSCVAT---LAKCSALTQKLPVGPPIANQVYILDSQLQVPVPG 4378
QY 1907 VMGLVVTGDLGARGYSK-ALDENRFVHTVNDQVTKAYRTGDRVRYRIGDGLIEFFGR 1965
Db 4379 VPGLHAGVGVARGVNLNPELTQKFIENPFGPS--KLYKTGDLGRY-LADGTIYELGR 4435
QY 1966 MDTQFIRGNRIRESIAEAALLRDSVRDAAVVLOQNEQDQAPFILGFVADVADHSDXK 2025
Db 4436 IDNOVKIRGFRIELGEIAVASOHPLVQESVVIAREDPCKRLVAYLVPAL-----QG 4489
QY 2026 Q-----SANQVEGWQDHFBESGMYSDIGEIDPSTIGSDFKGWTSMYDGGSDIDEMH 2076
Db 4490 QVLPEOLAQWQGVYSDWKLYEQSYSQOQTPDDPTF--NISGMNSSYTGKRAIPDSEMR 4547
QY 2077 EMIGETRTLLHDNRSLG--NVLEIGTSGMILNLDLSRLSYVGLPEPSRRAAFVFNKAT 2133
Db 4548 EWVESTVSR-----LCKQFQVLEICGSGLLFRVAPHCQYEWGADYSATIRNLERLC 4603
QY 2134 ESIPSLAGKAKVQVGTATDIGVDDLLPDLVNLNSYIQFSPSEYLAETAD-TLIHLPNV 2192
Db 4604 GETIQGLE-NVRLHKTADFEGIPQCAFDTVVNSVQVFPSPDYLLQVLEGAMTAIASQ 4662
QY 2193 QRIFFGDSQATNEHFLAARAIHTLGNKATKDDVRQKMAELEDMEELLVERPAFTSLK 2252
Db 4663 GKTFVGDVRSPLPLLPYHAQVLAARAESDKTVEQWQQQVHOTVAAEEELLIDPRFFIALQ 4722
QY 2253 DRPPGLVEHVEILPKMNEAVNELSAYRAYAVHVRGSLGDELVLPEK-DDWIDFQANOL 2311
Db 4723 QRP-QITWVEIQPKKGHQNELTQFYDVTLH-----LVLMGKSSLVKLATYQL 4773
QY 2312 NOKSLGDLKSSDAAIMAYSKFIETAFERQVVASLNSNIDWQ-----LSTI-----RS 2362
Db 4774 SMTLODLKLHVPELWLGIRDVFNORLOARQV-----LGMWENPAASTVGLERQRL 4826
QY 2363 SAEQSSLSVPDIFRITRAGAGFRYEVSSARQW---SQNGALDAVFHCCS-----2409
Db 4827 AAQPMYGINPEIQWQIGELGYTVHLS---WMQSSQDGCDFVVLRSNSTPVSVDIQSY 4882
QY 2410 ---QGRTLVNFPTDHLRGSDLLTNRLQRLQNRRIAREVRRLRSLPSYMPISNIV 2465
Db 4883 SFWEHETITTKPWTDY-----TNNPLYGLVKLVQVQVREFIOQKLPNYPQAFVL 4934
QY 2466 LDKMPLNANGKVDKRLSRAKVVVQKQATAA-PLPT---FPISEVEVILCEEATEVFGM- 2520
Db 4935 LNALPLTPNGKVDRRAL-----PTPDASRNLTGTFVSPRTPIEAQWAEVSEVLGLE 4987
QY 2521 KYDITDHPFNGLGSHLLATKLSRIDQRLKVRITVXDVFDHPFADLASVIRGLGLQ 2579
Db 4988 RIGVKDNFFELGHSLLATQVTSRSDIFSVLSLQNLLEYPTIVNLAQMI-EVLGVSQ 5045
```

RESULT 9

```
Q01135 PRELIMINARY; PRT; 5157 AA.
AC Q01135;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Peptide synthetase.
GN PESA.
OS Metarhizium anisopliae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Metarhizium.
OX NCBI_TaxID=5530;
```

```
RN SEQUENCE FROM N.A.
RP STRAIN=ME1;
RX MEDLINE=97082966; PubMed=8964498;
RA Bailey A.M., Kershaw M.J., Hunt B.A., Patterson I.C., Charnley A.K.,
RA Reynolds S.E., Clarkson J.M.;
RT "Cloning and sequence analysis of an intron containing domain of a
RT peptide synthetase from the entomopathogenic fungus Metarhizium
RT anisopliae.";
RL Gene 173:195-197(1996).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=ME1;
RA Bailey A.M., Reynolds S.E., Charnley A.K., Clarkson J.M.;
RT "Evidence for multiple peptide synthetases from Metarhizium
RT anisopliae.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; X89442; CAA61605.1; -
DR HSSP; P14687; 1AMU
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR003880; Pantne_attach.
DR Pfam; PF00501; AMP-binding; 4.
DR Pfam; PF00668; Condensation; 7.
DR Pfam; PF00550; pp-binding; 4.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS50075; ACP_DOMAIN; 4.
DR PROSITE; PS00455; AMP_BINDING; 4.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW Phosphopantetheine.
SQ SEQUENCE 5157 AA; 573954 MW; 1038242BA3143868 CRC64;
```

Query Match 12.5%; Score 2018.5; DB 3; Length 5157;
Best Local Similarity 23.9%; Pred. No. 4.9e-115;
Matches 795; Conservative 536; Mismatches 1278; Indels 711; Gaps 127;

```
QY 41 LDSSRTAEAKPCPTFPQDMLDCNALKDQSAIGHAVYDVPTDIDISRFALAKWEIVNQTPA 100
Db 2196 VKDCKDIEDYPCPTQEGLEMAISTRPRAYVARVARKIPSTIDLNRFRNAWQLISLEPI 2255
QY 101 LRAFATSDSGKTSQVILKDSFVSWMCWSSSSSPDEVV-RDEAAAAAGPCRNFRVILE 159
Db 2256 LRTRVLITNTNSV-QVVLREQ-----NPWKEGLTLESLKDNGLPIEYGLLHRLCIID 2309
QY 160 DMQTKKCOLWTFESHALVDVTFQVRSRFAAYKHEKDTHRPETPSSDATDTSQSVS 219
Db 2310 GL-GQDSYFIWSVHHALYDGSQNLFFERVYKQY-----MGDAISKSPSYNR 2355
QY 220 VYSMCEDNAVSAATHWOTH-LNDLNASVFPHLSHLMVNPNT--TTAEH--RITPPLSQ 274
Db 2356 FIHFLAQEDTKSANKFWKSOLLREQAPTSPP-----VLPSPYKPRADHVQTLRLPLSR 2409
QY 275 K---ALSNSAICRTALSILLSRYTHSDEALFG-AVTEQSLPFDKHYLADGYQTVAPLVR 330
Db 2410 QKHSTITSATLKAANAMVVSVQSRSDDIILALTLSGRTAPVPGIMKMTPTTITTVPLRI 2469
QY 331 HCSNLRASVDMAISSYDDRLGHLAPF---GLURDTRNTGDNGSAC-----DFQT 378
Db 2470 ELPS--RNTTVYKYLELVQNAQEMMPYEHVGTQSLRR-----LCLEAEPDYELDKH 2520
QY 379 VLLVTDGSHVNNINGFLQ---QITESSHFMPCNNRALLHCOMESSGALLVAYYDHNVI 435
Db 2521 LLVYQSLGEV---DGLFLELLPLEDDDF---NTYALIVQCSVEENVCIEARYDRDVI 2573
QY 436 DSLQTRTLRLLQQFGHLIKCIQSPDLSSMAEVNLMTEYDRAEIESWSQPLEVQDTLTHHE 495
Db 2574 ETQGMERMLWQFOHVLRLQNLNDESHSSTMGVDLVSPQDLQDLVWNSQKSHESMDICIQV 2633
QY 496 MLKAVSHSPTKT-AIQANDGDWTYSLELDNVSSRLAVHIIKSLGRQAQAIIPVYFEKSWV 554
Db 2634 IIEQMSRSPVDAICAWDGELSYQELDRLSNGLATHLVDLGV-GPEVMVPICFDKSAWT 2692
QY 555 IASMLAVLKSNGNAFTLIDPNDPPARTAAQVVYQTRATVALTSKLHRETQVOKLVGRCVVDD 614
```


1751 SLSYLGKDKTQIKHQRIEIGEIEHHV-LHCTKAVEVTVDVAVVPGEKNKSLVAFV-- 1807
QY 936 SSYFNGRPSDAHILD--HDATKAINIKLEQVLPRESIPSFYICMLPRTATGKIDRRRL 993
Db 1808 -----RSNGSTSTPQFYDNDPAIINELANSIPAYMIPTMYIQVPSIPTASGKTRKQL 1861
QY 994 RIMGKIIDKOTQGAIVOQAPAPVPFADTAALKHSIWQSLGIDPA-TVNVGAPFFELG 1052
Db 1862 REMGTAMASSHAARHWKHQNRPPV---TDMKHHVKLWARVLTLENAGEISLDDSFIRLG 1918
QY 1053 GNSITAIKWNAMARSVGMDLKVSNIOHPTLAG-----ISAVKGD-----LSVTLTPKSTH 1105
Db 1919 GDSIAAMKLVLSAAKAGLGLTAQIFRHTKLEDOQARHVTLTQGGPAPIAFSLPDDSPD 1978
QY 1106 -----EGPVEQSYS-----QGRLWFLDQDVGSLWYLIPIYAVMRGPVNDV 1146
Db 1979 VKALQADIRAYAEASSIEDYVCTPLQEGGLSLSSKPEYNTYTLQHVPELPTVDIQ 2038
QY 1147 ALRRALAALQORHETLRT-----FEDQGVQVIVHEKLSSEMKVVIDLGGSDLPDF 1198
Db 2039 QLRSAW-----EETIRTTDLTRIVLHPRYGL-VQVV---VKKEIQWHEPANADV--- 2085
QY 1199 EVLNQEOTTPFNLSSEACWRATLLRLGEDDHILITVMHHIISDGWSIDVLRDLNQLYSA 1258
Db 2086 -YIETDKQVQWVLGSLVRYAISPDTSASRKFIWTHALADGWTLDLILRKVKLAYST 2144
QY 1259 ALKDKDPLSALTPLPIQYSDFAK---QKDOFTEQKQLNWKKQKLDSSPAKIPTDF 1314
Db 2145 -----LHTVSPVS-EFRSVKVIITRTDMEV-----YKSTLGGYHSTTFPVC 2189
QY 1315 -ARPALLSGDAGCVHWTIDGELYQSLRAFPCNEHNTTSFVLLAAFAAHRYRTAVEDAVI 1373
Db 2190 VSTVAIEDSEVGQKH-----ELPRNITLSAHPST-----LLRAAWAIVQSNYSNTSDVVF 2240
QY 1374 GPIANRNR--PELEDIIIGCFVNTOCMKINIDHDTFGTLINQVKTATTAAPENEDIPE 1431
Db 2241 GEVFSGRSASVPFIEAIGVPMATLPVRKVIDDSELAREMLDLTLTOMIPHQOLGLO 2300
QY 1432 RV-----YSALQPSGKDLSSSTPLAOLIFAVHSOKDLGRFKFOGLBSVPVP 1476
Db 2301 RISQINTDQACASFOTLLVIQPPASTHNGQEPSLSFSGSPDYRLATYAL-GIECTPA- 2358
QY 1477 SKAYTRFDMEFHFOETSLGKSVNFADFLFKMETVENVRVFFELRNGLOSSRTPVSI 1536
Db 2359 SDGY-----SFCRARFDSKVLQAQAERMAOGLGHVVSOLVAVTASPSS 2404
QY 1537 LPLTDGIYV-----LEKLDVNLVKNHYDPRESS---LADVOTQVSAYPDSLAVVDSR 1588
Db 2405 TLVSIVLNTPODLEKLANWNAEVLGEQKHSMLLHQVFRKKALAAPQATAISSWDGE 2464
QY 1589 LTYTELDRQSDILAGLRRSRMPA---ETLVAFAPRSCETIVAFFGVLKANLAYLPLDV 1645
Db 2465 CSYAQLEKLSDALAAMLTDLGIGLDQOLVPLCFERSMWWVVMAMVLKTGAGIVPLDP 2524
QY 1646 RSPSARVQDILS--GLSGPTIVLGHDTAPPDIEVTVNEFVIRIDALNDSNADGFEVIEH 1703
Db 2525 AHPPSRHERILAKVIGG--CILVSPQAQROFGEGWTTW-----VSEASAAVPSI-H 2576
QY 1704 DSTKPSATSLA--YVLYTSGTGRPKGVMIHRVITRTV-TSGCIPNYPSETRMAHMATI 1760
Db 2577 AFDPTVTHLAVCWILFTSGSTGCEPKGYILEHGATCSYKLLGKTLGDKETRMHLHFSAY 2636
QY 1761 AFDGASYEISALLGRTLVCV-----DYMTTLDARALKDVFFREHVNAAASHVT 1809
Db 2637 AFDIATFIEIGLMSG-GCICIPSDAERLERLPQECTIF-----AVNTAILTP 2683
QY 1810 S-----SSQDVLPRVRLSRLLTMFFLVVTDSTAPDALDAQGLYQGVQ-CYNGYGPTE 1863
Db 2684 SVARLYTNDIP-----TLRSLCLA--GEAPNKQDISTWQHRIPLFCNYGPAAE 2731
QY 1864 GWMSTIYIPDSTESFINGVPIGRALNNSGAYVVDPE--QOLVIGVMGELVVTGGLARG 1921
Db 2732 ACLAATNIGPNADRSATRIGR-LRGVPLWITAPGNCRCKAPIGAVGELLIEGTLARG 2790

1922 YSDKALDENRFRVHIT-----VNDQTKAYRTGDRVRYRIGDGLIEFFGRMDTQF 1970
Db 2791 Y---WQLKPMHLHSRTSGCKCAQLGSDPADVDSTGPEIYKYDEDDGCVVYEGRKDNQV 2847
QY 1971 KIRGNRIESAETEAALLR-----DSSVRDAAVVL-----QONEDQAPEILG 2011
Db 2848 KIRGQRTELGELEYHLSQCFTPTAAEVVAVATSERDLASVTILVAFVKRETRDSSEKVA 2907
QY 2012 FVVAADHDSENDKGOSANOVGMOHFEESGMYSDIGEIDPS---TIGSDFGKWTSMYDG 2067
Db 2908 GIFALPSKLEHEINRRLLP-----YMPAVFVSYPEIPKPTATDKTDROKRLRELASVAT 2961
QY 2068 SOIDFDE-----MHE-----WL-----GETTRLHDN-----RSLGNV 2095
Db 2962 RAVDAPHQHPORLPSTVMEETLRDLWLKLVIPVQRTAIGLDSNFFRLGGDSIAALKLVGQA 3021
QY 2096 LEIG-----TGSGMI-----LFNLDLSRESV 2117
Db 3022 QOAGIELSKDIFLNPKLVDLAACCTDRRCVKEGSRMVAKHATISRFSLPLINASISIV 3081
QY 2118 -----GLEPSRSAAAF-VNKATESIPSLACK-----AKVOVGTATDIGOVDL--- 2159
Db 3082 DEVANACGIPPLVEDVPTCTPMQEGLSLSSRNFGCTVVSQIAIELAPDV--LVLDLFKLA 3139
QY 2160 -----HPDLVLVNSVIQ---YFPSSSEYLAETADTLIHLPNVQRIFFGD 2199
Db 3140 WQOTVSTWPIILSRIIQHPKGLGFLQAVLKEDVTWNNSTDLEVTDSSTP---MGFGS 3195
QY 2200 VRSQ-----ATNEHFLAARATH----- 2216
Db 3196 ELSRHALVMDNSGKHIFVYTVHHSIYDHTVLRLLDDVDYDNYKGNRERKDFPYTSFVRS 3255
QY 2217 -----TLGKNATKDD---VROKMAELEDMEELLVEPA----- 2246
Db 3256 VISMKSSSEEFWRNACKDEGSIIFPQRSLSRESCEDTIVVQSQVLTCTATGVTIMANVL 3315
QY 2247 -----FFTSCLKDR-----FPGILVEHVEILPKNM 2269
Db 3316 HAAMAVSSWHVGNQSIQVFTVLSGRTPAPVLGIENIAGPTIATAPPFVIIDPSETISNFS 3375
QY 2270 EAVNELSAYRYAAVHVVRGSLGDELVLVPEKDDWDIDFO-----ANQLNQSLGDLKAS 2322
Db 3376 PAIOGOMA---AVIHAQGLGQIRLSLSSACELACNFOTLFAVQEGRAMVGNLSGLKLL-- 3430
QY 2323 SDAATMAYSKIPFETAF---ERQVVASLNSN-IDEMOLSTI-----RSSABG 2366
Db 3431 -DVNTFSMTYALTDCFLDTGTFHVVKASFDSDRVVDWRMESILRQFGAVAQOLATKABG 3489
QY 2367 DSSLVSPDIFRIAGBAGFRVYSSARQWSONGALD--AVFHHCCSQ-----GRTLX-- 2415
Db 3490 GELVS---SIETLNEQGWEL---LRRMNSHRTKQWAVFPEDECKPSPIGAIGELLIEG 3542
QY 2416 -NFTPDHHLRGSDLLTNRLQRLQNRIRIAEVRER---LRSLLPSWIMPSINIVLDMKPL 2471
Db 3543 PDFF-----SKYLEDGARGVRPMDRNGHKTVLLTGLTILFADQNGNSHIHQKRT 3595
QY 2472 NANGKVDRKELSRRAKVPKQOATAAPLPTFPISEVEVILCEATE---VFGMKVDIOTH 2527
Db 3596 TISFGQORIDVSO-----IERHITSFLAGTEAVVEAIPAENSQSVLAVFLHRPELADR 3651
QY 2528 PFNLGGHSLATKLISRIDQRLKRVITVKVDHPFADLASVIRQGLGLOQVPSDGOQ 2587
Db 3652 GDNKSRPAICWSKQYDIEKNLSV-----VFPDMVPTL-----YIDMEAM 3691
QY 2588 DRSAB-----MAPRTETAILCDE-----FAKVLGFO---V 2615
Db 3692 PRTHGIDIDRSQLOTGLSLFPAEKVAIRLASRQKRPVATMOLAIRGLWASLIGAKEDTF 3751
QY 2616 GJTDFEGLGGHSLMATKLAVRIGHRDLTTVSVKQVDFHPVLFQALDNLNLSVQSKTNEI 2675
Db 3752 HLDDOFFFKSGGDSIGVIKL-VGEARKNTALAAADIFQPKLESIA-----VRATENTL 3804

Db 380 OLLSRVRLTQAYAHDPDFEFELVKVLPQ--ORDLSHTP-----LFQVMEVLQNAPI 430
QY 386 SHYNNINGFLOOITESSHMFNNRA---LLHCOMESSGALLVAYDHNVIDSLQTT 441
Db 431 SEVE-----LADLTISS--LPAQSATAFDLTSMQNTATGLVSWYENADLFDASTIE 482
QY 442 RLLOQFHLTK-CLOSPDLSSMAEVLNLTXYDRAE-IESWNSQPLEV-ODTLIHHMLK 498
Db 483 RMRGHFQTLLEGIVANPQE--QISQPLLTETEBQOQLLLKWNDDGVDPQDKCIHLLFEE 540
QY 499 AVSHSTKTAIQAWDGDWYSELNDVSSRLAVHIKSLGLRAQAQAIIPVFEKSKWIASM 558
Db 541 QVELTDAVAVYENQHLTYHELNCRANLAHLKSLGVKA-DVLVGICVERSLEMVVGL 599
QY 559 LAVLKSGNAFTLIDPNDPPARTAAQVVTQTRATVALTSKLHRETVQKLVGR-----C 609
Db 600 LGILKAGGAYPLASDYPTERLBLEDAQVSVLLTQ-----QKLIDRLPHTAQRVC 652
QY 610 VVDDDELQSVASDDFSSILTSKQSLAYVIFTSGTDPKGINIEHRAFSSCALKFGASL 669
Db 653 LDADWQVISOLS-QDNLIPEQTATNLAYVIYTSGSTQPKGVLIHAHGLLNLFVFWHOFTE 711
QY 670 GINSDFRALQFTHAFGACLELMTTLINGCCVICPSDDRMNSIPSFINRNVNMMMA- 728
Db 712 KITTLDKATQACTADAAWELWPLYTAGASIIYLVKPEILSPVDL-----QDWLESK 765
QY 729 --TPSYMGTFSPBDV-----PGLATVLVGEOMSSSVNATWAPKLOLLNGYGOSE-- 776
Db 766 KITISLPMPABOLLSLEWPESTTLRLMTGDKLHRYPSGLL--PFQVNNYGPTEW 823
QY 777 --SSSTCFASNSTENNNGRAVG-----AHSWIDPDINRLVPTGAVGELVIESPGIA 829
Db 824 VVSTGLVSVN--GRDNNISPPIGRPRIANVEIYILD--SYLQPVVPVGPCELHIGGAGIA 879
QY 830 RDYIVPPPEKSPFFTDIPSXP-----ANTFPDQAKLYRTGDLARYASDGS 876
Db 880 RGYLNRPOLTOEFKFIN-----PFKRSRGAGEORSRGETF-NSNRLYKIGDGLARYLPDGT 933
QY 877 IVCLGRIDSOVKIRGORVELGAIETHLRQOMPDDLIVVEATKRSOSANSTSLIAFLGS 936
Db 934 IEYLGRIQVQKIRGPRIELSETEAVLSQHSVDOLSCVI---AREDTGDKRLVAYV-- 988
QY 937 SYFGNRPDAHILHDHATKAINKLEQV---LPHRSIPSFYICMELPRTATGKIDRRR 992
Db 989 -----AH--QDCKPAIS--ELRQILKAKLPDYMVPFAFVLKSLPTNGKVDRA 1035
QY 993 LRIMGKDILDKOTQGAIVQQAIPVPFADTAAKLSHWQSIGIDPATVNVGATPFELG 1052
Db 1036 L---PKPDLDT---TILEKYVAPRTPPIEMLALL---NAQVLKLE--LVGIYDNEFFELG 1083
QY 1053 GNSITAIKMNVNARS-VGMOLKVSNIYQHTLAGISAVV-----KGDPLSYTLIPKSTHE 1106
Db 1084 GHSLLATQVRSIRTWFKVELPLRELFAFASATVVELAQSIGRLQOQDLELSTPFIQRAEN 1143
QY 1107 GPVEOSYSGRWLWFLDOLVGSLWLIPIYAVRMGPVNDALRALAALQORHETLRTTF 1166
Db 1144 AQULLSTAQORLWFLDQFEPNSALYNIPLALRLVGTNLQVLALEQSLHEIARHEALRTNF 1203
QY 1167 EQDQGVGVQIVHEKLSSEMKVIDL-----CGSDLOPFVELNQEQTPPFLNSSEAGWRATL 1221
Db 1204 VTVNGKPSQIQTQNTWTVSVVGLKHLSTTEQEIATQQLVROQAIEPFDLANQVLVRLT 1263
QY 1222 LRLEGDDHLLTIVMHHIISDGSVIDLRDLNOLISA-----ALKDSKDPL-----SAL 1270
Db 1264 LVLSETEHLLVCMHHIISDGSVMGVLOELAAALYNAYSQAQPSLRDATRSLLPRRGTP 1323
QY 1271 TPLTOYSDFAKWQDOFTE---QEKOLNWKOLKDDSPA--KIPTDFARPALLSGDAG 1325
Db 1324 APLTOYADFALWQR-QWLOGGVLOSQNLWYQOQLKD-APALLSPTDRPREAVOTFAGA 1381
QY 1326 CVHVTIDGLOSRAFNCNEHNTTFVLLAAAFRAAHYRLTAVEDAVIGTPTIANRRPEL 1385
Db 1382 HOEFALSQKLTQGLIOLISOKQVTLFMTLLAAAYDTLLRYRTGTEDILVGSPIANRRSEI 1441
QY 1386 EDIIGCFVNTQCMRINIDHDHDTGTLINOKVATTTAAAFENEDIPFERVVSALOPGSRDLS 1445
Db 1442 EGLIGFVNTLVWRTNLAGNPSFSELTUVRENAMEAYTYONLPPFEMLVALQP-QDRLS 1500
QY 1446 STPLAOLIFAVHSOKDLGRFKFOGLESVP--VPSKAYTRFDMEFHLFQETDSLKGSVNFAD 1504
Db 1501 HTELFQVMEALQN-VPLSGVELVGLRVTPLMPSERTAKFDNLFMQNTADGLGVWEYNT 1559
QY 1505 ELFKMETVENVVRVFEIIRNLNGLOSSRTPVSIILPLTDGIVTLEKLDLVANVHVVDYPRESS 1564
Db 1560 DLFDASTIERMTGHFVALLGIIIANPEQOISQISQPLLTVEHQHOLLWEMNDTGVYDQDLC 1619
QY 1565 LADVFOTQVSAYPDSLAVVDSSCRUTYTELDROSDILAGWLRRRSPAEFTLVAFAPRSC 1624
Db 1620 IYQLFEEQVORTPDVAVIFENQOOLYYQLNCRANOLANYLQSLGVSADVLVIGICVERSI 1679
QY 1625 ETIVAFPGVLKANLAYPLDVRSPARSARVODISGLSGPTIVILIGHDTAPPDIEVTNVFV 1684
Db 1680 EMVVGLLGILKAGGAVVPLDPDPYQORLSFILDAAALP--VLLTQOSLLESLEPHTAQVV 1737
QY 1685 RIRDALNDSNADGFEVIEHDSKPSAT-----SLAYVLTSGSTGRPKGVMIEHRIIRT 1739
Db 1738 CL-----SDSR---OLIEQHSRENPLTGSKPENLAYVIYTSGSTQPKGVQVSHNCVNF 1789
QY 1740 VTS-GCIPNYPSETRMAHMATIAFCGASYEISALLFRTLVCVDVMTTLD-ARALKDVF 1797
Db 1790 LASMHCPCGLTSDTFCAVTTISFDIAALELYPLIVGAKVVVVGVREVASDGAARLLEL- 1848
QY 1798 FREHVNAASHVTSSSODVP-----LRVPRRLSRTLMFFELVVTDTSTAPDALDAGLGVGVQ 1853
Db 1849 --QHSS-----TTVMOATPATWQMLVASGLSTQQLGKMLCCGGEALPPQLAHOLLETTGAQ 1901
QY 1854 CVNYGPTENGVMSTIPI--DSTE-----SFTNG-----VPIGRANNSGAYVVDPEQ 1900
Db 1902 VNNLYGPTETTWSICQLRNESTQLEARSATASGAALRAIPIGRPIANTQYIILDSHL 1961
QY 1901 QLVGIGVMGELVVTGDLARGYSK--ALDENRFVH-----ITVNDQTVK 1943
Db 1962 QPVPVAGELHIGGLGAKVYLRNRPOLTOEFKISPNFKRSRGAEEQSGRGETFNSD--R 2019
QY 1944 AYRTGDRVYRIGDGLIEFFGRMDTQFKIRGNRIESAEIEAALLRDSVDAVAVLQONE 2003
Db 2020 LYKTGDLARY-LPDGNIIEYLGRIDNQVKIRGRIELGEIEAVLSQHEHEDVQASCVIARQ-- 2076
QY 2004 DOAPEILGEVWADHDHSENDKQOSANQVEGWQDFHESGMYSOIGEIDPSTIGDFKGWTS 2063
Db 2077 ----- 2076
QY 2064 MYDGSOIDFEMHEWILGETTRTLHDNRSLGNVLEIGTSGMILFNLDLSRLESYVGLPSR 2123
Db 2077 -----DNGP-----DKRLVAY----- 2087
QY 2124 SAAAFVNKATESIPSLAGKAKVQVGTATDIGOVDDLLHPDLVVLVNSVIOVFPSEYLAETA 2183
Db 2088 ----- 2087
QY 2184 DTLIHLPNVORIFFGDRVSQATNEHFLAARAIHTLGKNATKDDVRQKMAELEDMEBELV 2243
Db 2088 ----- 2087
QY 2244 EPAFTTSLKDRFPGLVEHVEIILPKNMEAVNELSAYRYAAVHVHVRGSLGDELVLPEKDDW 2303
Db 2088 -----LVAHQDCKP----- 2096
QY 2304 IDFOANQLNOKSLGDLKSSDAAIMAVSKIPFEITAFERQVVASLNSNIDENQLSIRSS 2363
Db 2097 -----AIS----- 2099
QY 2364 AEGDSSLYPDIFRIAGEAGFRVEYSSARQWSQNGALDAVFHCCSQGRTLVNFPTDHL 2423
Db 2100 ----- 2099

QY 2424 RGSLLTNRLQRLQNRRIAIEVRERLRSLLPSYMPISNIVVLDKMPNANGKVDKELS 2483
DB 2100 -----ELRQFLAKLPEYWPVSAFVILESPLTPNGKVDRR--- 2135
QY 2484 RRAKVVVKQOATAPIPTPISEVEVILCEEATEVEFGMKVDITDHFENLGGHSELLATKLIS 2543
DB 2136 -----ALPK----- 2140
QY 2544 RIDORLKVRIITVKDFHPFADLASVIRQGLQOPVSDGOGQDRSAHMAPRTETEAL 2603
DB 2141 -----DLDTTILE-----KYVAPRTIEEML 2161
QY 2604 CDEFAKVLGFO-VGTDNFFDGLGSHLMATKIAVRIGHRLDTTVSKVDVDFHPVLQZAI 2662
DB 2162 ALLWAQVLKLELVGRDNEFFELGSHLLATQLVSRIRTKVFKVELPLRELFSTATFELAR 2221
QY 2663 ALDNLVQSTNEIVGGRMAEYSPQLLFTDPPEFMASEIKPOLQELIIDIYP-STQ 2721
DB 2222 STQ-----FQQQDLVLVSPILPRAENAE-----PLSYA 2252
QY 2722 MOKAFLEHHTTARPRF-VPEYIDFPSTSEPDAAGLIKACESLVNHLDFRTVFEAEAGE 2780
DB 2253 QORLNELOFENSALYINPMALRLVGLT--NOVALEQSLYEIIRHREALRTNFTVNGK 2310
QY 2781 LYQVVLSCLDLPQVI-----ETEDNINTATNEFLDEPAKEPVRIGH-PLIRFT--IHK 2831
DB 2311 PSQIITOTNNKVSVGLKHLSTTBOEI--ATQQLAQQAQIQPFDLANQALVRATLVLS 2368
QY 2832 QTKSMRVMIRSHALYDGLSEHVVRKLMHLYNGRS-----LLP-PROFSRYMOY---- 2880
DB 2369 ETKHI-LVLCMHVHISDGMVSFVLQELAALYNACSAQTSPLAPLPIQYADFALMQROW 2427
QY 2881 -TADGRESHGFRDVIQNTP--MTILSD-----DTVVDGNDATCKALHLKIVNIPSQ 2931
DB 2428 LOGDLQSLSTWQOOLKADAPALLSLPTDRPGPVQTFAGGHOEFALSVELSKNLTKLQ 2487
QY 2932 VLURGSNIITQATVFNACALVLSRESKDSKVYFGRIVSGROGLPVEYDQIVGPTCTNAVY 2991
DB 2488 E-QGCTLFMTLLAAFDT---LLRYVTGTFEDILVGSPIANRDRSEIE--GLIGFFVNTLV 2540
QY 2992 VRAHTESS--DYNOLLHDIDQVLLSLPHETIGFSLDKNCTDWPENAITNFCSCYIHNPE 3050
DB 2541 MRTNLNAGNFSFSELLGRVEMAEAYTQNLPEML-----VEALQ-----PHRDL 2587
QY 3051 YHP--ESQPEQORVEM-----GVLTKFVNIEMDEPLYDLATAGVEPDGAGLKVTIVIAK 3102
DB 2588 HTPLQVMPALQNVPLSEVELFGLSVTPPLMAESITAKFDLTLL--MONTGNGLVGWWEYN 2645
QY 3103 TOLFGKRKRVEHLLVEYSKTFEGL 3125
DB 2646 TOLFDASTIKRMGSHGVTLEGI 2668

RESULT 13

Q9RAH4
ID Q9RAH4 PRELIMINARY; PRT: 4379 AA.
AC Q9RAH4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Nosa.
GN Nosa.
OS Nostoc sp. GSV224.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=76334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GSV224;
RA Hoffmann D., Hevel J.M., Moore R.E.;
RT "Characterization of the nostoepetolide biosynthetic gene cluster of
RT Nostoc sp. GSV224.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF204805; AAF15891.2; -
DR HSSP; P14687; 1AMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00501; AMP-binding; 4.
DR Pfam; PF00668; Condensation; 4.
DR Pfam; PF00550; pp-binding; 4.
DR PROSITE; PS00075; ACP_DOMAIN; 4.
DR PROSITE; PS00455; AMP_BINDING; 4.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_3.
DR Phosphopantetheine.
SQ SEQUENCE 4379 AA; 491434 MW; 089F97BFC598E84 CRC64;

Query Match 12.2%; Score 1974; DB 2; Length 4379;
Best Local Similarity 21.5%; Pred. No. 2.1e-112;
Matches 847; Conservative 614; Mismatches 1294; Indels 1190; Gaps 130;

QY 82 IDISRFALAWKEIVNQTALPAFAFTSDSGKTSQVILKDSFVFSWCMWSSSSSSPDEVIRD 141
DB 43 LNTAAVNRVWEKIIIEHPILRT--TYTIEGKPVQOV--NOCHKFNLDVANASDWSEKQLKE 100

QY 142 EAAASAGP-----RCNRFVLLEDQTKKQLVWTFSHALVDVTFQORVLSRVFAA 192
DB 101 KIFAITDRPNLEKDSILRVNLFT-----RSKKEHILLTMMHIIAGDMWFDLLSEFQL 156

QY 193 YKHEKDTHRPEPESDSDATDTSQSVSVSMSCEDNAVSAT-----HEFWOTHLNDLNAS 246
DB 157 YAKEIEQVSEQEIQTAPDSITENKSY-LDFVRWQSEVLSASRGEKLMQYWKOL----AG 211

QY 247 VEPHLSHLMVNP-----TTTAEHRITFPLSQKALSNSA-----ICRTALSIILS 292
DB 212 ELPILNLFADKPRPPVQTYQGTYSKLDRLTIQKLYLALASGTSLYQVLLAAAFVLLY 271

QY 293 RYTHSDEALFGAVTEOSLPFDKHYLAD-----GTQTVAPLRVHCOSNLRASDVMDAIS- 346
DB 272 RYTNQTDILIGS-----PMRGWRGFEKIEIVFNVLNVLNVSVQENATFFKEFLAQVSK 325

QY 347 -----SYDRLGHLAPFGL-----RDIRNTGNGSAACDFQVLLVTDGSH----- 387
DB 326 TVREAQKHOD-----YFSLIARLOLPQDI-----SRSLPQO-----VSFTQWHRWCEP 371

QY 388 VNINGINGFTQDTESHSFPCNNRAL---LLHCOMESSGALLVAY-YDHNVIDLSQTLRL 443
DB 372 TENSLSHGGEQLLEIKRPLYLIGHGADFDLNLVMEAAQGVQLQLCWQYNTDLFEASTIRM 431

QY 444 LOQFGLHIKQO-----SPDLSSMAEVNLMTEVDRAE-IESWNSQPLE-VQDTLIHHEML 497
DB 432 A---GHFVTLLEGIVVNPQE--QIWQLPLLTETEQQOQLLVQWMDTGVLYLQNKCIHOLFE 486

QY 498 KAVSHSPTKTAOAWGDWITYSELDNVSSRLAVHIKSLGIRAOQAIPVYFEKSKWVIAS 557
DB 487 EQVDSTPNVAVVYENQQLTYQOLNRYRANQALHYLOSGLV-GENVLVGLCVLESLEWVG 545

QY 558 MIAVLKSGNAFTLIDNPDPARTAAQVVTOTRATVALTSKLHRTVOKLVGRCVVVDDELL 617
DB 546 LLGILKAGGAYVPLDPEYPAERLSEMLTDTHVVLTLTQERKLVESLPQHQAQVCLDLD-W 604

QY 618 QSVAS--DDFSLTKSQDLAVVIFTSGSGDPKGIHIEHRAFSSEKALFGASIGLSNDT 675
DB 605 HSISQANQDNLSTVSAENLSYVITSGSTGTPGVAVTHQAVNRVLVNTN-YIQTAD 663

QY 676 RALQFGTHAFGACILLEMTITLNG-----CVCIPSDDRMNSIPSEINRYNNVMMMA 728
DB 664 CVAQAANIAFDAATFEINGALLSGAKLIVITKSVLLLPPEFAVN-----LHKHKISVLF 718

QY 729 TP---SYMGTFFSPEDVPLATLVLEQMSSSVNAIW-----APKQLLLNGYGOSES 777
DB 719 TTALFNQLASVVPQAFSSRLYLLFGGE---AVDPKWLVEVLKGAPO-HLLHVGYPGTEN 773

QY 778 SSIC-----FASNMTSEPNNGRAV-GAHSWVIDPNDINRLVPTCAVCELVIESPGIA 829
DB 774 TTFSSWYLVEDLATTATTIP--IGRAIANTQIIVLDOS---LQVPVGPVGPGEIYLGAGLA 829

QY 830 RDYVPPPPKSPFFTDIPSWYPANTFP-----DGAKLYRTGDLARYAS 873
D 830 QGYFNRPELTKEF---IP-----NPFKAGSRREQBAGGRQKVEGSKLYKTGDLARYLA 881
QY 874 DGSIVCLGRIDSOVKTRGORVELCATETHLRQOMPDDLTIVVETATRSQSANSTSLIAEL 933
D 882 DGNIEYLGRIDNOVKTRGRFIEUSEIEAVLSQHDQVQSCVI---VREDTPGDKRLVAYV 938
QY 934 IGSSYFGNRPDAHILDHDATAKINIKLEOV-----LPRHSIPSYICMLPRTATGKID 989
D 939 VA-----HQNCKPTISELRQFLKAKLPDYWPVNAIVILESLPTPNKQVD 983
QY 990 RRLRIMGKIDILDKQGAIVQOAPAPIPVFADTAALKHSIWSQSLGIDPATVNVGATF 1049
D 984 RRALPT--PDLHSEQDKFVAPRNP-----IEEMLALLWQVLVRVE--LLGIDHNF 1031
QY 1050 ELGNSITALKMYNMARSV-GMDLKYSNIYQHFTLAGISAVVK-----GDPLSYTLIPKS 1103
D 1032 ELGGHSLLATQVLSRIRNVFKVLEPLRELFARATVAELAQLIROLOQOQDLESEAPILPK 1091
QY 1104 THEGPVEQSYSQGLWFLDQDVGSLWYLPYAVRMGPVNVDAALRAALAEQRHETLR 1163
D 1092 VRDAELPLSAQORLWFLDQLEPNSAFYNIPMALRLVGNLQNPALQESLLEIHRHEALR 1151
QY 1164 TTEDODQGVQVHEKLS-----EMKVTDLCGSLDPPE-----VLNQEQTTP 1208
D 1152 TNPITVDGKPSQIOTRREQGTNRQGTVSIVDF--KDLSTNEQETASKQLARQQAIEP 1209
QY 1209 FNLSSBAGWRATLLRGEDDHILTIVMHHIISDGWSIDVLRRLDNLNOLYSAALKSDKPLS 1268
D 1210 FDLANGALITLLVISETEHILLVCMHHVDSGMSGVFQVQELAALYDAYAQCP--S 1266
QY 1269 ALTPLPQYSDFAKQKD--QFIEQEKQLNYWKQKQKDS-SPAKIPTDFARPALLSGDAG 1325
D 1267 SLAPLPQYADFAIQREWLGQVNLQNLTYWQOQLANAPSLSLTDRPRPAVQTFAGA 1326
QY 1326 CVHVTIDGELYQSLRAFCEHNTTSFVLLAAFAAHYRLTAVEDAVIGTPIANRNP 1385
D 1327 YQFALSASLTLGTLSQKQGTLEMTLAAFDTLRYTGAEDILVGSPIANRDRSEI 1386
QY 1386 EDIIGFVNTQCMRINIDHDTGTILINOVKATTAAFENEDIPFVVSALQPGSRDLS 1445
D 1387 EGLIGFVNTLVMTDLAGNPSSELLTRVRETAEMAYSHOHLPFEMLVEALQ-ERDLS 1445
QY 1446 STPLAQILFAVHOKDLGRKFQGLE-SVPVPSKAYTRFDMFHLFQETDLSKGSVNFAD 1504
D 1446 HTPLFQVMEFLQN-APISTFELNQLSVSPTLEGTTAKFDLTLSMENTATGVVGVWEYST 1504
QY 1505 ELFKMETVENVRVFFELRNGLOSRTPVSIILPTDGIIVLEKDLVNLVHVDYDPRESS 1564
D 1505 DLFADASTIERMTGHFVTLLESIIANPQOSISQSLPLLVEQOQLLVENNNQVQDYPLDKC 1564
QY 1565 LADVFOQVSAYPDSLAIVDSSCLRTVTELDROSDILAGWLRRRSPAPETLVAFAPRSC 1624
D 1565 LHQLFEEQVALTADAVAVFENQOLTYQQLNCRANELAHYLSQGLVEADVGVICVERSL 1624
QY 1625 ETIVAFPGVLKANLAPLIDVRSPARVQDILSCLSGPTILVICHDTAPDIEVNTVEFV 1684
D 1625 LMVVGLLGILKAGGAVVLPDPDPQERLSFMLED-AQLRVLLTQH----- 1668
QY 1685 RIRDALNSNAD-----GFEVTEHDS-----TKPSATSLAYLYVTSGSTGRKPGVMIEH 1733
D 1669 QLKEKLPQHOQGVVCLDTMQFISQSSQENLIITVQASNLAYIYVTSGSTGRKPGKAMNTH 1728
QY 1734 RVI-----IRVTSCINPSETPRMAHMATIAFDGASYETIVSALLFGRTLCVCD 1783
D 1729 LGICNRLNMQAYQILTALDCI-----LOKTPFSDVSVWEEFWPLITCARLVAK 1779
QY 1784 YMTTLARALKDYFFREHVNAAHSVTSQQ-----DWPLRVPRLSLRMTLFFLVVTDSTA 1839
D 1780 PGGHKDSAYLVNLILEQVTHVHFVPSMLQVLFEEQNLENCRSLKR-----VICSGEAL 1833

QY 1840 PDALDAQGLYQGVQC--YNGYGPTEGVMSYIIPIDSTESFING----VPIGRALNNSGA 1893
D 1834 PVELQ-ERFARFARCELHNLGYPTAAIDTYW-----QCFPNGLRTPVIGRAIANTQI 1887
QY 1894 YVVDPEQQLVIGVMGELVYTGDLARGYSDKA-LDENREV-----HIT 1936
D 1888 YILDEHLQVPVGVAGELHFIAGLAKGYLNRRLPDTTKDFIPNPFREVGEQSGSKAKIL 1947
QY 1937 VNDQTV-----KAYRTGDRVRYRIGDGLIEFGRMDTQFKIRGNRIESAETALLRDSV 1992
D 1948 PMSQSLVPNPOLYKTGDLARY-LPDGTIEYIGRIDNOVKIRGFRIELGELEVLNQCEDV 2006
QY 1993 RDAVVVLQONQOAPILGEFVADHDHSE----- 2021
D 2007 QASCVIAREGTTGDKCLVAYVW-PHQHSKPTTNELROFLKAKLPDYMPVSAFVILESMPL 2065
QY 2022 -----NDK----- 2024
D 2066 TSNKGYVDRRALPAPDLSSEIKDKYVAPRTPDIEMLAQTWSQVLOLPOVGHNDHFFELGGH 2125
QY 2025 -----GQSANQVEGWQ----- 2035
D 2126 SILLATQVSRIRNIFKVELPLRELFATAALAEAREIEQLQQQDLQLSAPPLPRAENAQ 2185
QY 2036 -----DHEF----- 2039
D 2186 LALSAYAORLWFLDOFEPNSPFYNNMSIALRAGTVEVALEQSFQOIIHRHEALRTNFT 2245
QY 2040 ----- 2039
D 2246 VDCQATQIIOTETNWTIVSIVDLKNLSTTKOEIASOQLVQOQATOPPELATQALVRATLV 2305
QY 2040 -----SGMYSDIGEIDPSTIG-----SDFKGWTSM 2064
D 2306 LSETEHILLMCIHHIVSDGWSIGLFLQELAAALYNAYSQGRSSTLAPLIOYADFALWQR 2365
QY 2065 YDGSQIDFDEMHEWL-----ETTYTLNDRSL-----GNVLEIG 2099
D 2366 WLQGEVQSLQSWQOKLASAPITLLPTDRPRPSVQYQGTGHOEFALSSELTGKAKKS 2425
QY 2100 TGSQMLIF-----NLDORLESYVGLPSPSAAAFVNKATESIPSLAGAKAKVQGTATD-- 2152
D 2426 QEOGVTLFWTLAAFDTLRYTGTEDILVGSPIANRDRSEIEGLGFFVNTLVMTDLS 2485
QY 2153 -----TGQVDDL-----HPDLVVLSNVIQVFPSS-- 2176
D 2486 GNPFSSELLGRVEMAMEAYSHOHLPFEMLVEALQPERDLSHSPLFQVMEVQLQNAPTSGL 2545
QY 2177 -----EY----- 2178
D 2546 ELTGLTVSSLPKIGTTSRFDLTIMONSPGLIGVWEYNTDLPDASTIERMTGHFVTLLE 2605
QY 2179 -----LAEIADTLIHL----- 2190
D 2606 GIITANPEERISQPLLTOLPOOKLLVWENDTQVDYQDLCIHLFEKQVDSPTDAITALIA 2665
QY 2191 -NVORIFFGVRSQATNEHFLAARAIHTL-----GKNATKDDVQRMAE----- 2233
D 2666 RSANAVVFEDQOLTYVELNYRANQLAHYLQSLGVSQDALVGLCVERSLEHVIIGLLILKA 2725
QY 2234 -----LEDMEELIVEPAFTSLKDRFPGLVHEVEILPKNMEAVNE 2274
D 2726 GAAYVPLDPEYPPERLSFLEDAQVSVLLTQ-----OSILDRLOHQANQVCLDQDAQLIS 2782
QY 2275 LS-----AYRAAVHVHVRSLG-----DELVLPEKDDWIDFQANOLNQ 2313
D 2783 CSODNLISDVQANNLAYIYVTSSTGPKGIAMNOLALSAILWHRENKXIPRGAKTLQF 2842
QY 2314 KS-----LGDLLKSSDAAIM-----AVSK--IPFE----- 2336
D 2843 ASINFDVFOEIFTWCSSGTLFLIGELRRDTSALLGFLQQAIRMFPLFPVALQOLAE 2902
QY 2337 -----ITAFERQVVA-----SUNSN 2351


```
Db 2903 VAIGELVNSHLREIITAGEQLQITPAISQWLSKLTCTLHNNHGPSSESHLATSPTLTNS 2962
QY 2352 IDEWQ-----LSTIRSAEGDSSLS
Db 2963 VETWPLLPVGRPTANAQIYILDRFLQPVGVGPGELYIAGVLLSQGFNRPELTLEKFI 3022
QY 2373 PDIF---RIAGEAGFRVSS-----ARWSON-----G 2398
Db 3023 PNPFKSRGAGEQSGRGETFNCDRLYKTGDLARYLSDGNIEYLGRIQNOVKIRGFRIELG 3082
QY 2399 ALDAVFHCCSQGRTLNVNPTDHLRGSDLTNRPL-----QRLQNRRIAEVRELRGS 2452
Db 3083 EIEAVLSQDQVASCAMATPAAGIAR-EDIPGNKRLVAYIVPQKEQKLTVPF-VRSFLKS 3140
QY 2453 LLPSYMIPSNVIVLDMKPLNANGKVDKELSRRAKVPKQQTAAPLTPFPPISEVEVILCE 2512
Db 3141 KLPEYVPSAIVILDALPLTPNGKLDLR-----ALPAPDL----- 3175
QY 2513 EATEVFGMKVDITDHFNLGGHSLATKLISRIDQRLKVRITVKDVPDHPFADLASVIR 2572
Db 3176 -----HSQLLDK----- 3182
QY 2573 QGLGLQOPVSDGQDRSAHMAPRTETAILCDEFKVLGPF-QVCGITDNFEDLGHSLMA 2631
Db 3183 -----YVAPRNPIBEILSLIWAQVLKVEQVGIHDFNFFELGGHSLIA 3223
QY 2632 TKLAVRIGHRLDVTTSVKDFHPVLFOLATALDNLVOSKNEIVGGREMAEYSPFQLLF 2691
Db 3224 TQLISRVTSKVELPLSLFAAPVAIVELS---QNIQSQOQDL----- 3264
QY 2692 TEDPEEFMASEIKPOLELOEITQDIYPTQOMKAEFLDHTTARPPFPVFFYIDFSTSEB 2751
Db 3265 -----ELIASPLTPRAENAE-----PLSYAQTRCLWFLDKLPNSAFYNFIALRLVGT 3314
QY 2752 DAAGLIKACESLVNHLDFRTVFAEASGELVQVLSCLDLPLOVETEDNINTATNEFL- 2810
Db 3315 NRTALEQSLQEIIRHEALRTNFVIVDGKPSQIIQITQINCTLSVVDFK-HLSTTQETAS 3373
QY 2811 DEFAPKPEYRLGH-PLIRFTTIKOTKSMR-VIMRISHALYDGLSLHVVYKRLHMLYNGRS- 2867
Db 3374 QQLVQKQPDLANEALVRATLVLSLSETHALLVCMHHIVSDGSMGVFVQELAAALNAYSQ 3433
QY 2868 -----LPL-PRQFSRYMOY-----TADGRESGHGFWRDVIQNTPTMIL-----SDDT 2908
Db 3434 GEPSPLTPLPIQYADFATQWRNLQGVQVLSQVWQHLANAPALLSLPTDRPRPSMOT 3493
QY 2909 VVDGNDATCKALHLS-KIVNIPSOVLRGSSNIITQATVFNACALVLSRESKDVVGR 2967
Db 3494 FAGAYQOPALSKELDKLTQLSQO--QGVTLFMTLLAAPDT-----LLRYTGTEDILVGS 3547
QY 2968 IVSGRQGLPVEYODIVGCTNAPVVRHIESS-DYNQLLHDIDQOYLLSLPHETIGFSDL 3026
Db 3548 PIANDRSEIE--GLIGFLFNTVMRTNLAGNPSSELLGRVREMAIDAYSHQNLFPFEM 3605
QY 3027 -----KNCNTDWPAINFNSCITYHNFEYHPESQFQQRVEMGLVTKFVNIEMDEPLY 3080
Db 3606 VEALQPERDLSHTP-----LFQVMFVLQN-----APMSQOELTGLTVSPVVKSTTKF 3655
QY 3081 DLAIAGEVPEPGAGLVKVIYIAKTLQFGRKRVRHLLLEYSKTFEGL 3125
Db 3656 DUTLG--MENTATGLVGIWEYNTDLFDASTIERMTGHFLTLEGI 3698
```

RESULT 14

```
Q44928
ID Q44928 PRELIMINARY; PRT: 4450 AA.
AC Q44928;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Gramicidin S synthetase 2.
GN GRS2.
```

```
OS Bacillus brevis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Paenibacillaceae; Brevibacillus.
OX NCBI_TaxID=1393;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NAGANO;
RX MEDLINE=95122465; PubMed=7822255;
RA Saito F., Hori K., Kanda M., Kurotsu T., Saito Y.;
RT "Entire nucleotide sequence for Bacillus brevis Nagano grs2 gene
RT encoding gramicidin S synthetase 2; a multifunctional peptide
RT synthetase.";
RL J. Biochem. 116:357-367(1994).
DR EMBL; D29676; BAA06146.1; -
DR HSSP; P14687; IAMU.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR000379; Ser_estrs_site.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00501; AMP-binding; 4.
DR Pfam; PF00568; Condensation; 4.
DR Pfam; PF00550; pp-binding; 4.
DR Pfam; PF00975; Thioesterase; 1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00075; ACP_DOMAIN; 4.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_2.
DR PROSITE; PS00455; AMP_BINDING; 4.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_4.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_4.
SQ SEQUENCE 4450 AA; 508674 MW; F3197E77BF69316D CRC64;
Query Match 12.0%; Score 1933; DB 2; Length 4450;
Best Local Similarity 22.0%; Pred. No. 7.7e-110;
Matches 845; Conservative 587; Mismatches 1306; Indels 1098; Gaps 137;
QY 67 KQSAIGHAVYDVPTDIDISRFALAKKEIVN---QTPALRAFAFTSDSGKTS----- 114
Db 19 QEGMLFHALLDKNAHLVQMSIATGIVDVLLSESLNILDYDVFRTTFLHEKIKQP 78
QY 115 -QVILKDSFVSMCWSSSSSPDEVVDEAAAA---ASGPRCNRFVLLD-----MOT 163
Db 79 LQVILKERPV--OLOFKDISLDEEKREQAISOYKYODGETV--FOLTRPLMRVAIFOT 134
QY 164 KKC--QLVMTFSHALVDVTFQQRVLSRVAAYKHEKDHTRPETPESDADTDTDSQSVSV 221
Db 135 GKVNQVIMSFHILLMDGWCNFIIFNDLFIYLSKE---KKPLQLEAVQPKYKIKWL 190
QY 222 SMSCEDNAVSAATHFWQTHLDNLNASVPHLSHLMVNPNTT---TAHRITFP----- 271
Db 191 E---KQDKQALRYNKEHLMNYDQSV--TLPKKAAINNTTVEPAQFPAFQKVLQQLL 245
QY 272 --LSQKALSNAICRTALSILLSRYTHSDEALFGAVTEQSLPDKHYLAD--GTYQTVP 327
Db 246 RIANOSQVTLNIVFQITWIGVILQKYNSTNDVVGVSVV--SGRPFSEISGIEKMVGFLTNTLP 304
QY 328 LRHVQCSNLRASDVMDAISS---YDDRGLHGLAPFLGRDIRNTGDNGSAACDFOTVL---- 380
Db 305 LRIQTKQDSFIELYKTVHQNVLFSQOHEY---FPLYEQNHTELQNLIDHIMVIENYP 361
QY 381 LVTDGSHVNGINGFLQO-----ITESHFMPCNNRALLHQMESSGALL---VAYYDHN 433
Db 362 LVEELQK-----NSIMQKVGFTVRDYKMFEPNTNYDMTVMLVPRDEISVRLDYNAAYDID 416
QY 434 VIDSLQTTLLQOFGHLIK---CL-QSPDLSSMAEVNLMTEVDRAE--TESNNSQPLEVQ 488
Db 417 FTRKIE-----GHMKEVALCVANNPHVL---VQDVPLTTKQEKHLLVELHDSITEP 466
QY 489 DTLIHHEMLKAVSHSPTKTAIQAWGDGWTYSSELDNVSSRLAVHKSIGLRAQAIIIPVVF 548
Db 467 DKTIHQLFTEQVEKTPPEHVAVVFEDEKVTYRELHERSNQLARLREKGVK--KESIIGIMM 525
QY 549 EKSXVVIASMLAVLSGNAFTLIDNPDPARTAAQVVTQTRATVALTSKLHRETQVKLVGR 608
```


QY 161 -----MOTKKCOLWTFSHALVDTFOQVRLSRVFAAYKHEKDH-----RPEPPES 207
D 130 LLRFTALRSADRVRVMTNHHIIVLDGWSMPVLLRELMAALYAAEGDPSALPRVRYDYL 189
QY 208 S--DATDTSQSVSVMSCEDNAVSATHFWQTHLNDLNASVPFPHLSHLMVFN--PTTT 263
D 190 AWDARD-----AARDWRRSLSGLEAT-----LLAPDGPAST 226
QY 264 AEHRITPLSQ-----KALSNSAICRTALSILLSRYTHSDLEALFGAVTE--Q 308
D 227 APSQVSTVDSVSGALSAMARGCVTNTVVGANALALAAQATGRDDVVFGATVSGRPP 286
QY 309 SLFPDKHYLAGTYQTVAPLVRHCQSNLRASDYMDAISYDDR--LGLAPFGLDIRNTG 367
D 287 ELPQVESMI--GLFINTPLVRARLDQAEPLGDLFRRLQNEQARLLDHPW--GLADIQHW 343
QY 368 DNGSACDFQTVLLVDTGSHVNGINGFLQIITESHFMPCNNRALLHCOMESSGALLV 427
D 344 GHGEL---FDTAMVFN-----YFVEGDLTAPADPDL-----RVASADIKGG 384
QY 428 AYDHNVIDSLQTRL-----LOQFG--HLIKLOSPLDLSSMAEVLNMT 471
D 385 THEAVNVATMGAELSFRVDYRPDLDEAYARDFGRMLRVLET-----LISD 433
QY 472 YDR--ABIES-----WNSQPLEVQDTLIIHEMLKAVSHSPTKTAIQAWGDWT 517
D 434 PDRPVAHLDTLPAVRERVLVEVNGAPTQLPGTPLHELISEQARLTDPDAVAVCDGTSLT 493
QY 518 YSELNDVSSRLAVHKSGLRAQOALIPVYFEKSKWVIAASMLAVLKSNAFTLIDNDPP 577
D 494 YALDGRANLQARHLERLUGAED--FVAIALHSLDAVMTSLAVLMTGAAYLPIDPDYA 552
QY 578 ARTAQVVTOPRATVALTSKL-----HKEVQKLVGRVGVVDDDELQSVASDDFSLSLK 631
D 553 ERITYMLDARPALTLREPVPAAAYGHRPT-----DD--VTDAERTPWAL-- 597
QY 632 SODLAVIFTSSTGDPKGMIEHRAFSKALK-----FGASLGINSOTRALQGTGTAFGA 687
D 598 --HAAYMIYTSSTGRPKGVVIEHHALATVHLRARTYNTAMTV-----TVLHSPALFADL 650
QY 688 CLEIMTTLINGCVCIPSDDDRMNSIPSPINRYNNMMATPSYM-----GTFSPED 740
D 651 TITALTPTLGTAGTVHLTSLEE--AEVQPSLIK-----ATPSHLPLTNLAATASPSH 701
QY 741 VPGIATLVVGEQSSSVNAIWA---PKQLLNGYQSESS--SICFASNMSPE--PNNMG 794
D 702 -----TLILGGEALHTDQLTDRWTHQPGAOIINAYGPTSTVNIITDRLDGTGEEGPVIG 756
QY 795 RA--VGAHSWIDPNDINRLVPI--GAVGELVIESPGIARDYIVPPPEKSPFTDIPSWY 851
D 757 RPFANTQVYVLD-----SALRPVAPGTTGELYLAGELQARGYLGRPALTAERETAN----- 807
QY 852 PAMTFPDGAKLYRTGLARYASDGSIVCLGRIDSQVKIRGORVELGAIETHLRQOMPDDL 911
D 808 PHSSVP--GARMYTGLAHNHHGHTYDGRADHQIKLGRHREPEIETTLAQ----- 861
QY 912 TIIVETAT--KRSOSANSTSLIAFLIGSS--YFGRNRPDSAHILDHATKAINIKLEQVLP 968
D 862 TGITOATVOLREDPQDQRLVAVLVTSYDENTVRDA-----LVSALPDY 907
QY 969 SIPSFYICMLELPTATGKIDRRRLIMKIDLDKOTGAIVQOAPAPVPF--ADTAAK-- 1026
D 908 MVSALVTLDALPTNGKLDRTAL-----PAPYASASTTGR 945
QY 1027 -----LHSIWQSLIGDPATVNVNGATFFELGGNSITAIKMVMNAR--SVGMCLKVSN 1077
D 946 PRTPREILCTLYSEILSVN--TVGIDDSFFDLGGHSLATRLVSRIRTLGRELPIRL 1003
QY 1078 YOHPTLAGISAVVKGDPLSYTLIPKSTHEGPEVQSYSGRLMFLDOLDVGSWLYLIPYAV 1137
D 1004 FETPTVAGLSRALDTSGLTALTARPRERIPSLPSAQORLWFLHOLEGPTATYNIPTTL 1063

QY 1138 RMGRPVNVADLRALAALEQRHETLRTTFEDQDQGVQIV--HEKLSSEMKVIDLCSGDL 1195
D 1064 RLGTGLTDALOSAFNDLLARHESLRTTYTDDQGAQIVLPVEAVATPPAVVDVAAEDV 1123
QY 1196 DPEVYNQEQTTTFENLSSEAGWRATLLRLGEDDHLIVMHHIISDCWSIDVLRDLNQL 1255
D 1124 --AERVAEAAAHAFDLGAETIPVARLFRVSEREHVLLLVHHTIASDAMSGRPLAQDLTAA 1181
QY 1256 YSAALKDKSDPLSALTPLPIQYSDFAKWKQDQFTEQ-----KOLNYMKKOLKD--SSP 1307
D 1182 YTARCADD--APAWQPLPVQYADYALWQOEIILGDDTDPDTLAGRQAYWKQOAGLPEQ 1238
QY 1308 AKIPTDFARPALLSGDAGCVHVTIDGELYQSLRAFCEHNTTSEVLLAALAFRAHYRLTA 1367
D 1239 LDLPDTPRPAATADHTGDRVTFVPADLTRLTELARETNTTAFMWIQAALALLTRHGA 1298
QY 1368 VEDAVIGTPTANRRPELEDIICGFVNTQCMRINIDHHDFTGFLINOVKATTAFAFED 1427
D 1299 GEDIPGTVPAGRTDDATDHLIGFFVNTLVLTDTSGNPTFRDILLTRVRDTLTAYTHQD 1358
QY 1428 IPEVVVSALQPGSRDLSSSTPLAQIFAVHSOKDLGRFKOGLESVPVPSKAYT--RFDME 1486
D 1359 LPERLVEALNP--TRSLTHPLFQTMLTJLHNTOGTKDRFAGLAAEVVASESVSARDLS 1417
QY 1487 FHLQFETDS-----LKGSYNFADELFKMETENVVRFFEILRNGLQSSRTVPSILPT 1540
D 1418 FALAHEFGADHSCAGSGGVTYSTALFDRATVRDLADRLVRLAAVAHAPGRSVQGLEIM 1477
QY 1541 DGIVTLEKLDVNVKHVDYPRESLADVQFQVSAYPDSLAVDSSCLRYTTELDRSDI 1600
D 1478 DAERLVLQEWENDTAAE--PPAASVTGLFERQARRSPGATALEFGEVRLSTAEULANAR 1536
QY 1601 LAGWLRRRSPAPETLVAVFAPRSCETIVAFPGVLKANLAYLPDLDVRSRPSARVQDILSGLS 1660
D 1537 LARHLVARGAAPGRFVAVALPRSAELVVTLLAVLKSAAVLPIDPHYPADRVEVHLAD-A 1595
QY 1661 GPTIVILIGHTDAPDIEVTNVFVRINDALSNADGFEVIEHDSKPSATSLAYLYTS 1720
D 1596 GPALTV---TEP---VAEAGLSGYGDA--DLGADELRGPVHGAHP-----AYTIYS 1639
QY 1721 GSTGRPKGVMIHRVIRTVTSCIPNY-----PSETRMAHMTATFAGASVEY 1770
D 1640 GSTGRPKGVV-----VPRGALDNFLADMGRRTPGSGDRLLAVTTVGFDIAGLEIF 1690
QY 1771 SALLFGRTLVCVDYMTTLDAARLKDVFREHVNAAASHVT-----SSSQDV 1815
D 1691 LPLHGAVLVLADEETARDPHALL-----HRVSASGITMVQATPSLWQGVAAVAGDELA 1744
QY 1816 PLRV-----PRLSRTLMFFFLVVTDTAPDALOAGLYQGVQCYNGYGPTEGVMST 1868
D 1745 GVRVLVGGEALPSELARAL-----TD-----RARSVTNLYGPTTEATIWAT 1784
QY 1869 IYPIDSTESGFINGVPIGRALNNSGAYVVDPEOQLVGIVMGELVVTGDLGARGYSDK-AL 1927
D 1785 --AADVAES---GPVIGRPLANTSAVLDALRPVGVPGELYLAGELQAGYHILRPAL 1839
QY 1928 DENRFVHTIIVNDQTVKAYRTGDRVRYRIGDGLIEFFGRMDTQPKIRGNRTIESAIEAAL 1987
D 1840 TSERTADPYGPAGTRMYRTGDLV--CRRRDGALRYLSRVDOQVKLRGRIELGEIEAELS 1898
QY 1988 RUSSVDRDAVVLQOONEDQAEIILGFVVADHDHSENKQOSANQOVGQDHFESGMYSIDG 2047
D 1899 RHPVAEASV-----RHPVAEASV----- 1908
QY 2048 EIDPSTIGSDFKWTSMYDGSQIDFDEMHEWLGETTTLHDNRSLGNVLGIGTSGMILF 2107
D 1909 -----IVREDRPG----- 1916
QY 2108 NLDRLSLESYGLPEPSRAAFAVFNKATESIPSLAGAKAVQVGTATDIGQVDDLHPDLVLN 2167
D 1917 --DRRLVGVV----- 1924
QY 2168 SVIQYPSSEYLAETADTLIHLPNVQRIFGDVRQATNEHFLLAARAIHTLGNKATKDDV 2227

```
Db 1925 ----- 1924
QY 2228 RQMAELEDMEELVERAFTSLKDRPGLVEHVEILPKNMEAVNELSAYRYAAVVHR 2287
Db 1925 ----- 1930
QY 2288 GSLGDELVPVEKDDWIDFOANQKNOKSLGDLKSSDAALMAVSKIPFEITAFERQVVAS 2347
Db 1931 ----- 1930
QY 2348 LNSNIDENQOLSTIRSAEGDSSLVDPDIFRIAGAGFRVEVSSARQWSONGALDAVFHHC 2407
Db 1931 ----- 1943
QY 2408 CSQGRTLNFPDHHLRGSDLLTNRQLRQNRRTAIEVRERLRSLLPSYMTIPSNIVVLD 2467
Db 1944 ----- 1966
QY 2468 KMLPNANGKVDKELSRRAKVVVKQOATAPLTFPISEVEVILCEEATEVFGMKVDITDH 2527
Db 1967 ----- 1986
QY 2528 FENLGGHLLATKLISRIDQRLKVRITVVDVDFHVPFADLASVIRQGLQOPVSDGOGQ 2587
Db 1987 ----- 1991
QY 2588 DRSAHMAPRTETEAILCDEFKVLGF-QVGIITDNFFDLGGHSLMATKLAVRIGHRLDTTV 2646
Db 1992 ----- 2047
QY 2647 SVKQVDFHVLQALADNLVQSKTNEIVGGREMAEYSPFOLLETPDEEFMASEIKPQ 2706
Db 2048 ----- 2082
QY 2707 LELQELIQDIYSTOMQKAFLEDHTARPRPVFPYIDFPSTSEPDAGLIKAC-ESLVN 2765
Db 2083 ----- 2133
QY 2766 HLDIFRTVFAEASGELYOVWLSC-----LD-LPIQVETEDNINTATNEFLDEFKPEV 2818
Db 2134 ----- 2192
QY 2819 RLGHPLIRFTTIKQKSMRVIMRISHALYDGLSLEHVVRKHLMYNGR--SLLP-----P 2871
Db 2193 ----- 2246
QY 2872 HQFSRYMQYTAD--GRESGH-----GFWRDVIQNTPTMTILSDDTVVVDGNDATCKALH 2921
Db 2247 ----- 2303
QY 2922 LSKIVNIPSOVLRGSSNIITQAT-----VFNAACALVLSRESKDVVFGRIVSGROGL 2975
Db 2304 ----- 2363
QY 2976 PVEYODIVGCPCTNAVVPVRAHIESS-DYNOLLHDIOQYLLSLPHETIGFSDL-----KR 3028
Db 2364 ----- 2421
QY 3029 NCTDWPEAITNFSCCITYHNFYHPESQEQVEMGVLTKFVNITEMDEP----- 3078
Db 2422 ----- 2463
QY 3079 -----LYDLAIA-GE-VEPDG--AGLKVTVIATKTOFGKRKRVEHLEEV 3118
Db 2464 NATRMTAGTGVAKFDLALTFGRHDPGGPGAGMRGSLRYTELFDRETAESVISRL 2519
```

Search completed: May 30, 2003, 12:48:34

Job time : 245 secs

THIS PAGE BLANK (USPTO)